

Result No.	Score	Query Match	Length	DB	ID	Description
1	1332	100.0	638	19	AAW76391	Pseudomonas aerugi
2	1323	100.0	652	14	AAK36808	Pseudomonas Exotox
3	1323	100.0	652	14	AAK32456	PE with Inactivator
4	1323	100.0	668	14	AAK32457	PE having M1 residu
5	1323	100.0	746	14	AAK36810	Full-length PE with
6	1318	99.6	746	14	AAK32458	PE having M1 residu
7	1310	99.0	668	14	AAK36809	Full-length PE with
8	1207	91.2	937	14	AAK36822	PE binding/transloc
9	1207	91.2	937	14	AAK32470	PE binding and trans
10	1195	90.3	414	14	AAK36807	Pseudomonas exotox
11	1195	90.3	414	14	AAK32455	PE amino acids 2-4

12	1195	90.3	426	14	AAR36806	PE domain1 and
13	1195	90.3	426	14	AAR32454	PE(2-414)-Ma(57-68
14	1195	90.3	613	14	AAR40102	Pseudomonas exotox
15	1195	90.3	613	14	AAR40113	Pseudomonas exotox
16	1195	90.3	613	19	AAW59278	Pseudomonas aerugi
17	1195	90.3	613	20	AAW62915	Pseudomonas sp., ex
18	1195	90.3	613	20	AAW62912	Pseudomonas sp., ex
19	1195	90.3	613	21	AAW67462	Amino acid sequenc
20	1195	90.3	614	16	AAW87738	Native pseudomonas
21	1195	90.3	665	14	AAW36805	Pseudomonas exotox
22	1195	90.3	665	14	AAAR32453	PE(2-414)-M1(2-252
23	1195	90.3	670	14	AAAR36820	Pe-Influenza A vire
24	1195	90.3	617	14	AAAR32468	BSPe1c5a fragmen
25	1195	90.3	917	14	AAAR36821	Pe binding/transid
26	1195	90.3	917	14	AAAR32469	Pe binding and t
27	1190	89.9	613	14	AAAR40105	Pseudomonas exotox
28	1190	89.9	613	14	AAAR40106	Pseudomonas exotox
29	1190	89.9	613	14	AAAR40107	Pseudomonas exotox
30	1190	89.9	613	14	AAAR40108	Pseudomonas exotox
31	1190	89.9	613	14	AAAR40110	Pseudomonas exotox
32	1190	89.9	613	14	AAAR40111	Pseudomonas exotox
33	1187	89.7	613	14	AAAR40104	Pseudomonas exotox
34	1187	89.7	613	14	AAAR40109	Pseudomonas exotox
35	1187	89.7	613	14	AAAR40112	Pseudomonas exotox
36	1181	89.3	634	20	AAV43547	Aa mutant chimeric
37	1181	89.3	634	20	AAV43545	A chimERIC toxin c
38	128	9.7	25	21	AAW67463	Amino acid sequenc
39	93.5	7.1	1684	18	AAW25671	HA6C3 protein. Ho
40	93.5	7.1	1684	19	AAW46761	Amino acid sequenc
41	93.5	7.1	1704	19	AAW46771	Amino acid sequenc
42	91	6.9	16	9	AAW34349	Peptide related to
43	88	6.7	315	19	AAW56261	Mature interleukin
44	88	6.7	359	19	AAW56260	Construct containin
45	88	6.7	380	21	AAW5236	IL-13 binding chain

```

XX PF 24-MAR-1998; 98WO-US05710.
XX XX
XX PR 26-MAR-1997; 97US-0042056.
XX XX
XX PA (TEXA ) UNIV TEXAS SYSTEM.
XX XX
XX PI Draper RK;
XX XX
XX DR WPI: 1998-532023/45.
XX DR N-PSDB: AAV61784.
XX PT Delivering compounds to cells as new conjugate with detoxified
XX PT exotoxin A - able to cross membranes and deliver to the cytoplasm,
XX PT e.g. nucleic acids, antibodies, tumour suppressors etc.
XX PS Disclosure: Page 58-60; 76pp: English.
XX XX
XX CC This is the amino acid sequence of Pseudomonas aeruginosa
XX CC exotoxin A (ETA), a virulence factor and secreted protein encoded
XX CC by the toxA gene (see AAV61784). The invention employs a group of
XX CC membrane penetrating proteins, of which ETA is an example, to
XX CC carry a variety of therapeutic agents across the cellular membrane
XX CC and into the cytoplasm. Such therapeutic agents may be a nucleic
XX CC acid, peptide, peptide nucleic acid, (single chain) antibody, or
XX CC other pharmaceutical. The nucleic acid is particularly DNA under
XX CC control of a eukaryotic promoter (e.g. the cytomegalovirus
XX CC immediate-early or beta-actin promoters) encoding a nucleic acid
XX CC binding protein, single chain antibody, tumour suppressor,
XX CC cytokine, hormone, toxin, or is an oncogene. Alternatively, it
XX CC encodes an antisense molecule that targets an oncogene or viral
XX CC protein. ETA can be modified to suppress its toxicity while
XX CC retaining its ability to transport molecules across membranes, or
XX CC by introduction of a free cysteine in domain III (see AAV76393) of
XX CC ETA to permit branched coupling and the production of non-terminal
XX CC conjugates of ETA, specifically peptide nucleic acids (see AAV61787).
XX CC The methods allow decreased doses of therapeutic agents to be
XX CC administered, thereby reducing toxicity, and provide effective
XX CC targeting of an agent to a specific point within a cell.
XX XX
XX SQ Sequence 638 AA;
XX
Query Match 100.0%; Score 1323; DB 19; Length 638;
Best Local Similarity 100.0%; Pred. No. 2.2e-129;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

XX XX
XX DT 25-AUG-1993 (first entry)
XX XX
XX DE Pseudomonas Exotoxin with inactivated toxin domain.
XX DE
XX XX Vaccine; cytotoxic T lymphocyte; CTL; influenza A virus;
XX KW matrix protein; Ma; cell recognition domain; translocation domain;
XX KW ADP-ribosylating domain; anti-viral agent.
XX XX
XX OS Pseudomonas aeruginosa.
XX XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 650
XX FT /note= "corresponds to nonsense codon TAA"
XX XX
XX PN EP541335-A.
XX XX
XX PD 12-MAY-1993.
XX XX
XX PF 04-NOV-1992; 92EP-0310067.
XX XX
XX PR 08-NOV-1991; 91US-0792507.
XX XX
XX PA (MERI ) MERCK & CO INC.
XX PI Donnelly JJ, Friedman A, Howe LA, Liu MA, Marshall MS;
XX PI Montgomery DL, Oliff AA, Shi X, Ulmer J;
XX DR WPI: 1993-154266/19.
XX DR N-PSDB: AAQ41719.
XX XX
XX PT Recombinant DNA encoding bacterial toxin-antigen conjugates - are
XX PT useful as vaccines against viral infections, tumours and
XX PT parasites
XX XX
XX PS Example 7; Page 34-37; 81pp: English.
XX XX
XX CC Plasmid pVCPE/2 was made by replacing the 105bp PuuM1/EcoRI
XX CC fragment of pVC45DpF+ (containing the Pseudomonas Exotoxin coding
XX CC region) with a 46bp DNA fragment (AAQ41718) encoding an in-frame
XX CC duplication of PE codons 604 to 613 flanked by unique restriction
XX CC sites. This construct is used for generating full-length molecules
XX CC of PE with the deletion of residue 553 resulting in an inactivated
XX CC toxin domain (AAQ41719) fused to chosen protein segments between
XX CC PE codons 604 and 605.
XX CC
XX XX Sequence 652 AA;
XX
Query Match 100.0%; Score 1323; DB 14; Length 652;
Best Local Similarity 100.0%; Pred. No. 2.3e-129;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db	121	swslnwlvpighekpnsikvfihelnagngqlshmspiytlmgdellaklardatffvra	180			
Qy	181	HESNEMQPTLAIASHAGVSVVMAQTQPRREKRWSEASGKVLCLLDPLDGVYNYLAOQRN	240			
Db	181	hesnemqptlaishagvsvvmaqtqprrekrwseasgkvlclldpldgynylaqqrn	240			
Qy	241	LDOTWEGKIYRV	252			
Db	241	lddtwegkiyrv	252			
RESULT 4						
AAR32457						
ID	AAR32457 standard; Protein; 668 AA.					
XX	AAR32457;					
XX						
DT	20-JUL-1993 (first entry)					
XX						
DE	PE having M1 residues 55 through 67 between residues 604 and 605.					
XX						
KW	PE; Pseudomonas exotoxin; Influenza A virus; M1; matrix protein;					
KW	fusion; hybrid: pVCPE/2-Na; pVC-PEM1-2; ompA; signal sequence;					
KW	promoter; ribosome binding site; RBS.					
XX						
OS	Synthetic.					
XX						
FH	Key Location/Qualifiers					
FT	MSC-difference 666					
FT	/note= "residue not defined in the specification;					
FT	encoded by stop codon TAA"					
XX						
PN	EP532090-A.					
XX						
PD	17-MAR-1993.					
XX						
PF	02-SEP-1992; 92EP-0202660.					
XX						
PR	09-SEP-1991; 91US-0756249.					
XX						
XX	(MERI ) MERCK & CO INC.					
PA						
XX						
PI	Donnelly JJ, Friedman A, Hawe LA, Liu MA, Marshall MS;					
PI	Montgomery DL, Oliff AI, Shi X, Ulmer J;					
XX						
DR	WPI: 1993-087107/11.					
DR	P-PSDB; AAR32457.					
XX						
PT	Bacterial toxin-antigen protein conjugates - to elicit cytotoxic					
PT	T-lymphocyte immune response, used for preventing viral					
PT	infections, e.g. by Influenza virus, HIV and human					
PT	papilloma:virus					
XX						
PS	Disclosure; Page 40-42; 85pp; English.					
XX						
CC	Example 8 describes the construction of pVCPE/2-Ma.					
CC	pVCPE/2-Ma was made by ligating into the XmaI site of pVCPE/2 a 48					
CC	base pair DNA fragment encoding amino acids 55 through 67 (AAQ36020).					
CC	This construct expresses in E.coli full-length PE with M1 amino acids					
CC	55 through 67 inserted between PE amino acids 604 and 605 (AAQ38394).					
CC	One may replace the ompA signal sequence with the promoter/ribosome					
CC	binding site as described for pVC-PEM1-2.					
XX						
SQ	Sequence 668 AA;					
Query Match 100.0%; Score 1323; DB 14; Length 668;						
Best Local Similarity 100.0%; Pred. No. 2.3e-129;						
Matches 252; Conservative 0; Mismatches 0; Indels 0; Caps						
Qy	1	MHLPIHPWIPVASLGLLAGSSASAAEAEDLMNECAKACVLDLKDGVRSRMSVDPDPAIA	60			
Db	1	mhlpihpwipvaslqlagssasaaeeafdlmnecaacvldlkdavrsrmsvdpdpaia	60			

QY 61 DTNGQGVLYHSWVLEGGNDALKLAIDNALSDGLTIRLEGGVEPNKPVRYSTRQARG 120  
 Db 61 dtngqgvlyhyswvleggndalklaidnalsdgltrleggvepnkpvrystrqarg 120  
 QY 121 SWSLNLVPIGHEKPSNLIKVFHIELNAGNLSHMSPIYTIEMGDELLAKLARDATFFVRA 180  
 Db 121 swslnlvpi ghekp snlikvf hielnagnls hmspiyti emgde lla klard atffvra 180  
 QY 181 HESNEMQPTLAISHAGVSVVMAQTQPRREKRWSEWASGKVICLLDPLDGVVNYLAQORCN 240  
 Db 181 hesnemqptlaishagvsvvmaqtqprrekrwsewasgkviclldpldgvvnylaqrcn 240  
 QY 241 LDDTWECKIYRV 252  
 Db 241 lddtweckiyrv 252

RESULT 5  
 AAR36810  
 ID AAR36810 standard; Protein; 746 AA.  
 XX AC AAR36810;  
 DT 25-AUG-1993 (first entry)  
 XX DE Full-length PE with Influenza A virus M1 codons 15-106 inserted.  
 XX KW Vaccine; cytotoxic T lymphocyte; CTL; influenza A virus;  
 KW matrix protein; anti-viral agent; Pseudomonas exotoxin.  
 XX OS Chimeric Pseudomonas aeruginosa.  
 OS Chimeric Influenza A virus.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 744  
 FT /note= "corresponds to nonsense codon"  
 XX EP541335-A.  
 XX 12-MAY-1993.  
 XX 04-NOV-1992; 92EP-0310067.  
 XX 08-NOV-1991; 91US-0792507.  
 XX (MERI ) MERCK & CO INC.  
 XX Donnelly JJ, Friedman A, Hawe LA, Liu MA, Marshall MS;  
 PI Montgomery DL, Oliff AA, Shi X, Ulmer J;  
 XX WPI: 1993-154266/19.  
 DR N-PSDB; AAQ41724.  
 XX Recombinant DNA encoding bacterial toxin-antigen conjugates - are  
 PT useful as vaccines against viral infections, tumours and  
 PT parasites  
 XX Example 9; Page 45-47; 81pp; English.  
 XX Plasmid pVCPE/1-M1:15-106 was made by subcloning a PCR-amplified  
 CC DNA fragment encoding M1 amino acids 15 to 106 into the XmaI site  
 CC of pVCPE/2 (see AAQ41718). The primers used in the amplification  
 CC reaction had sequences AAQ41722 and AAQ41723. The construct expresses  
 CC in E.coli full-length PE with M1 amino acids 15-106 inserted  
 CC between PE amino acid 604 and 605 (see AAQ41724).  
 XX Sequence 746 AA;  
 SQ

Query Match 100.0%; Score 1323; DB 14; Length 746;  
 Best Local Similarity 100.0%; Pred. NO. 2.8e-129;  
 Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHLIPHWPILVASLGLLAGSSASAAEEAFDLWNECAKACVLDLKDGVSRSSMSVDPATA 60  
 Db 1 mhliphwplvaslgl lagssasaaeeafdlwneca kacvldlkdgvsrssmsvdpata 60  
 QY 61 DTNGQGVLYHSWVLEGGNDALKLAIDNALSDGLTIRLEGGVEPNKPVRYSTRQARG 120  
 Db 61 dtngqgvlyhyswvleggndalklaidnalsdgltrleggvepnkpvrystrqarg 120  
 QY 121 SWSLNLVPIGHEKPSNLIKVFHIELNAGNLSHMSPIYTIEMGDELLAKLARDATFFVRA 180  
 Db 121 swslnlvpi ghekp snlikvf hielnagnls hmspiyti emgde lla klard atffvra 180  
 QY 181 HESNEMQPTLAISHAGVSVVMAQTQPRREKRWSEWASGKVICLLDPLDGVVNYLAQORCN 240  
 Db 181 hesnemqptlaishagvsvvmaqtqprrekrwsewasgkviclldpldgvvnylaqrcn 240  
 QY 241 LDDTWECKIYRV 252  
 Db 241 lddtweckiyrv 252

RESULT 6  
 AAR32458  
 ID AAR32458 standard; Protein; 746 AA.  
 XX AC AAR32458;  
 DT 20-JUL-1993 (first entry)  
 XX DE PE having M1 residues 15 through 106 between residues 604 and 605.  
 XX KW PE; Pseudomonas exotoxin; influenza A virus; M1; matrix protein;  
 KW fusion; hybrid; pVCPE/2-M1:15-106; pVCPE/2; ompA; signal sequence;  
 KW promoter; ribosome binding site; RBS; primer; PCR; amplification.  
 XX OS Synthetic.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 744  
 FT /note= "residue not defined in the specification;  
 FT encoded by stop codon TAA"  
 XX EP532090-A.  
 XX 17-MAR-1993.  
 XX 02-SEP-1992; 92EP-0202660.  
 XX 09-SEP-1991; 91US-0756249.  
 XX (MERI ) MERCK & CO INC.  
 XX Donnelly JJ, Friedman A, Hawe LA, Liu MA, Marshall MS;  
 PI Montgomery DL, Oliff AI, Shi X, Ulmer J;  
 XX WPI: 1993-087107/11.  
 DR N-PSDB; AAQ38397.  
 XX Bacterial toxin-antigen protein conjugates - to elicit cytotoxic  
 PT T-lymphocyte immune response, used for preventing viral  
 PT infections, e.g. by influenza virus, HIV and human  
 PT papilloma: virus  
 XX Disclosure; Page 48-50; 85pp; English.  
 XX Example 9 describes the construction of pVCPE/2-M1:15-106.  
 CC pVCPE/2-M1:15-106 was made by subcloning a PCR-amplified DNA  
 CC fragment encoding M1 amino acids 15 through 106 into the XmaI site  
 CC of pVCPE/2. The sequence of the oligonucleotide primers used to  
 CC amplify the M1 segment are those shown in AAQ38395-96, respectively.  
 CC This construct expresses in E.coli full length PE with M1 amino acids  
 CC 15 through 106 inserted between PE amino acids 604 and 605 (AAQ38397).

CC One may replace the ompa signal sequence with the promoter/ribosome  
CC binding site as described for pVC-PEM1-2.  
XX  
SQ Sequence 746 AA:

Query Match 99.6%; Score 1318; DB 14; Length 746;  
Best Local Similarity 99.6%; Pred. No. 9,2e-129;  
Matches 251; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHLLPFWITVLVSLGLAGSSASAAEAFDLMECAKACVLDKGVSSRMKSVDPATA 60  
DB 1 mlllpfwitvlvaslglagssasaaeeafdlmecaacvldkgyvssrmvpaia 60  
QY 61 DTNGGCVLHYSWVLEGGNDALKAIDNALSTSDGLTIRLEGGVEPNKPVRYSYTROARG 120  
DB 61 dtnggcvlhysswvleggnalakaidsalstsdglitirlegvpepnkpvrysytrqarg 120  
QY 121 SWSLNLVPIGHEKPSNIXYFTHELNAGNOLSHMSPIYTIEMGDELLAKIARDATFFVRA 180  
DB 121 swslnlvpiqhekpnsikxfthelnagnolshmspiytiemgdeliakardatffvra 180  
QY 181 HESNEMOPTLAISHAGSVVMAOTOPRREKRMSEASGVLCULDPDGYVNYLAQORCN 240  
DB 181 hesnemoptlaishagsvvmaqtprekrwsewsgkvlclldpdyvnylaqqrn 240  
QY 241 LDDTWEKGIYRV 252  
DB 241 lddtwegkiyrv 252

## RESULT 7

AAR36809  
ID AAR36809 standard; Protein: 668 AA.

AC AAR36809;

DT 25-AUG-1993 (first entry)

XX Full-length PE with Influenza virus M1 fragment inserted.

KW Vaccine: cytotoxic T lymphocyte; CTL; Influenza A virus;

KW matrix protein; M2; cell recognition domain; translocation domain;

KW ADP-ribosylating domain; anti-viral agent.

XX Chimeric Pseudomonas aeruginosa.

OS Chimeric Influenza A Virus.

XX Key Location/Qualifiers

FT Misc-difference 666

XX EP541335-A.

PD 12-MAY-1993.

PF 04-NOV-1992; 92EP-0310067.

PR 08-NOV-1991; 91US-0792507.

XX (MERI ) MERCK & CO INC.

PI Donnelly JF, Friedman A, Howe LA, Liu MA, Marshall MS;

PI Montgomery DL, Oliff AA, Shi X, Ulmer J;

DR WPI; 1993-154266/19.

DR N-PSDB; AAQ41721.

XX Recombinant DNA encoding bacterial toxin-antigen conjugates - are

PT useful as vaccines against viral infections, tumours and

PT parasites

PS Example 8; Page 36-42; 81pp; English.

XX Plasmid pVCPE/2-Ma was made by ligating into the XmaI site of  
CC pVCPE/2 (see AAQ41718) a 48bp DNA fragment encoding amino acids 55 to  
CC 67 of Influenza A virus. The resulting construct expresses in E.coli  
CC full-length PE with M1 amino acids 55-67 inserted between PE amino  
CC acid 604 and 605 (see AAQ41721).

XX Sequence 668 AA:

Query Match 99.0%; Score 1310; DB 14; Length 668;  
Best Local Similarity 99.6%; Pred. No. 5.3e-128;  
Matches 251; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHLLPFWITVLVSLGLAGSSASAAEAFDLMECAKACVLDKGVSSRMKSVDPATA 60  
DB 1 mlllpfwitvlvaslglagssasaaeeafdlmecaacvldkgyvssrmvpaia 60  
QY 61 DTNGGCVLHYSWVLEGGNDALKAIDNALSTSDGLTIRLEGGVEPNKPVRYSYTROARG 120  
DB 61 dtnggcvlhysswvleggnalakaidsalstsdglitirlegvpepnkpvrysytrqarg 120  
QY 121 SWSLNLVPIGHEKPSNIXYFTHELNAGNOLSHMSPIYTIEMGDELLAKIARDATFFVRA 180  
DB 121 swslnlvpiqhekpnsikxfthelnagnolshmspiytiemgdeliakardatffvra 180  
QY 181 HESNEMOPTLAISHAGSVVMAOTOPRREKRMSEASGVLCULDPDGYVNYLAQORCN 240  
DB 181 hesnemoptlaishagsvvmaqtprekrwsewsgkvlclldpdyvnylaqqrn 240  
QY 241 LDDTWEKGIYRV 252  
DB 241 lddtwegkiyrv 252

## RESULT 8

AAR36822  
ID AAR36822 standard; Protein: 937 AA.

AC AAR36822;

DT 25-AUG-1993 (first entry)

XX PE binding/translocation domains-HIV gag fusion protein.

KW Vaccine: cytotoxic T lymphocyte; CTL; Human Immunodeficiency Virus;

KW AIDS; anti-viral agent; Pseudomonas exotoxin; fusion construct.

XX Chimeric Human Immunodeficiency Virus.

OS Chimeric Pseudomonas aeruginosa.

XX EP541335-A.

PD 12-MAY-1993.

PF 04-NOV-1992; 92EP-0310067.

PR 08-NOV-1991; 91US-0792507.

XX (MERI ) MERCK & CO INC.

PI Donnelly JF, Friedman A, Howe LA, Liu MA, Marshall MS;

PI Montgomery DL, Oliff AA, Shi X, Ulmer J;

DR WPI; 1993-154266/19.

DR N-PSDB; AAQ41731.

XX Recombinant DNA encoding bacterial toxin-antigen conjugates - are

PT useful as vaccines against viral infections, tumours and

PT parasites

PS Example 26; Page 70-74; 81pp; English.

CC A fragment containing the HIV gag gene was obtained from plasmid  
 CC HIVBR322 by PCR with primers that added a SacII site adjacent to  
 CC the Arg codon of gag (to give the sequence AAQ41729), and a SacI site  
 CC immediately after the termination codon at the 3' end (to give the  
 CC sequence AAQ41730). The PCR fragment was cloned into plasmid  
 CC pV45DF+7 (containing Pseudomonas Exotoxin sequence). In the  
 CC resulting plasmid, pVc-ompa-PEGAG, the binding and translocation  
 CC domains of PE were fused to the gag gene of HIV-1. The fusion  
 CC contains an ompa leader.

XX Sequence 937 AA:

Query Match 91.2%; Score 1207; DB 14; Length 937;  
 Best Local Similarity 95.5%; Pred. No. 5e-117;  
 Matches 231; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 11 VASGLIAGGSSAAAEAFDLWNECAKACVLDLKDGVSSRMSVDPALADTNGGVLHY 70  
 DB 10 valagfatvagaanaaeeafdlwnecaacvldlkdyvrssrmvdpaladtnggvlyhy 69  
 OY 71 SWLEGGNDALKLAIDNALSTSDGLTIRLEGGVEPNKPVRYSTTRQARGSWSLMWLVPI 130  
 DB 70 smvlegndalklaidnalsitsdgltrlegvpepnkpvrysttrqargswslmwlvpi 129  
 OY 131 GHEKPSNIVFTHLHNAQOLSHMSPTITIEGDELLAKLARDATFFVRAHESNEMOPLT 190  
 DB 130 ghekp snivfthelnaqolshmsptitiegdel laklar datffvrah esnemoplt 189  
 OY 191 AISHAGSVYMAQTOPRREKRWSEWASGKVLCLDPLDGYYNYLAQRCNLDDTWEKGIY 250  
 DB 190 aishagsvymaqtgprrekrwsewasgkvlclldpldgynyylaqrcnlddtwegkly 249  
 OY 251 RV 252  
 DB 250 rv 251

# RESULT 9

AAR32470 AAR32470 standard; Protein; 937 AA.

XX AAR32470;

DT 20-JUL-1993 (first entry)

DE PE binding and translocation domains - HIV GAG fusion protein.

XX PE: Pseudomonas exotoxin; Influenza A virus; M1; matrix protein;

KW fusion; hybrid; pVc-ompa-PEGAG; HIV-1; HIVBR322; pVc45DF+T;

KM PCR; amplification; binding; translocation; domain;

XX leader sequence.

PN EP532090-A.

PD 17-MAR-1993.

PF 02-SEP-1992; 92EP-0202660.

PR 09-SEP-1991; 91US-0756249.

XX (MERI ) MERCK & CO INC.

PI Donnelly JF, Friedman A, Howe LA, Liu MA, Marshall MS;

DR Montgomery DL, Oliff AI, Shi X, Ulmer J;

XX WPI; 1993-087107/11.

PT Bacterial toxin-antigen protein conjugates - to elicit cytotoxic  
 PT T-lymphocyte immune response, used for preventing viral  
 PT infections, e.g. by influenza virus, HIV and human  
 PT papilloma virus

XX Disclosure; Page 73-77; 85pp; English.

PS Example 26 describes the construction of pVc-ompa-PEGAG.  
 XX The HIV GAG gene was obtained from plasmid HIVBR322 by  
 CC PCR with oligonucleotide primers which added a SacII site  
 CC adjacent to the Arg codon of GAG to give the sequence of  
 CC AAQ38412, and a SacI site immediately after the termination  
 CC codon at the 3' end to give the sequence shown in AAQ38413.  
 CC The PCR fragment was digested with SacII and ligated to the  
 CC plasmid pVc45DF+T, which had been digested with EcoRI, the 5'  
 CC overhang filled in by Klenow fragment, and digested with SacII.  
 CC The resulting plasmid was named pVc-ompa-PEGAG (AAQ38414) and was  
 CC verified by a partial sequence at the SacII junction. This  
 CC construction fuses the binding and translocation domains of PE  
 CC to the GAG gene of HIV-1 virus. The fusion protein contains an  
 CC ompa leader sequence. Alternatively, any vector contg. the complete  
 CC coding region for HIV GAG can be used with these oligomers to generate  
 CC the HIV GAG gene by HIV.

XX Sequence 937 AA:

Query Match 91.2%; Score 1207; DB 14; Length 937;  
 Best Local Similarity 95.5%; Pred. No. 5e-117;  
 Matches 231; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 11 VASGLIAGGSSAAAEAFDLWNECAKACVLDLKDGVSSRMSVDPALADTNGGVLHY 70  
 DB 10 valagfatvagaanaaeeafdlwnecaacvldlkdyvrssrmvdpaladtnggvlyhy 69  
 OY 71 SWLEGGNDALKLAIDNALSTSDGLTIRLEGGVEPNKPVRYSTTRQARGSWSLMWLVPI 130  
 DB 70 smvlegndalklaidnalsitsdgltrlegvpepnkpvrysttrqargswslmwlvpi 129  
 OY 131 GHEKPSNIVFTHLHNAQOLSHMSPTITIEGDELLAKLARDATFFVRAHESNEMOPLT 190  
 DB 130 ghekp snivfthelnaqolshmsptitiegdel laklar datffvrah esnemoplt 189  
 OY 191 AISHAGSVYMAQTOPRREKRWSEWASGKVLCLDPLDGYYNYLAQRCNLDDTWEKGIY 250  
 DB 190 aishagsvymaqtgprrekrwsewasgkvlclldpldgynyylaqrcnlddtwegkly 249  
 OY 251 RV 252  
 DB 250 rv 251

# RESULT 10

AAR36807 AAR36807 standard; Protein; 414 AA.

XX AAR36807;

DT 25-AUG-1993 (first entry)

DE Pseudomonas exotoxin domains I and II encoded by pVc-PEBR.

XX Vaccine; cytotoxic T lymphocyte; CTL; Influenza A virus;

KW matrix protein; Ma; Pseudomonas exotoxin; cell recognition domain;

KM translocation domain; anti-viral agent.

XX Pseudomonas aeruginosa.

OS EP541335-A.

PN 12-MAY-1993.

PD 04-NOV-1992; 92EP-0310067.

PR 08-NOV-1991; 91US-0792507.

XX (MERI ) MERCK & CO INC.

XX Donnelly JJ, Friedman A, Howe LA, Liu MA, Marshall MS;  
 PI Montgomery DL, Oloff AI, Shi X, Ulmer J;  
 XX WPI: 1993-154266/19.  
 DR N-PSDB: AAQ41715.  
 XX  
 XX Recombinant DNA encoding bacterial toxin-antigen conjugates - are  
 PT useful as vaccines against viral infections, tumours and  
 PT parasites  
 XX  
 XX Example 5: Page 30-32; 81pp; English.  
 PS  
 XX Control plasmid pVC-PEB7 encodes a T7 promoter-driven gene fusion  
 CC consisting of PE amino acids 2-414 followed by termination codons,  
 CC instead of by at least part of the influenza A virus Matrix  
 CC protein (as in e.g. AAQ1714).  
 CC  
 XX  
 SQ Sequence 414 AA:

Query Match 90.3%; Score 1195; DB 14; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-116;  
 Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 AEEAFDLMNECAKACVLDKDGVRSSRMVDPADITNGGVLYHSMVLEGNDALKLAI 85  
 |||||||  
 DB 2 aeafdlwneccakacvldkdyvrssrmvdpaditnggvlyhsmvlegndalkl 61  
 OY 86 DNALSTSDGLTIRLGGVEPNKPVRYSTYRQAGSLSMLVPIGHEKSNIKVFIHEL 145  
 |||||||  
 DB 62 dna1stsdgltrlrgvgnkpvrystyrqagswslmvlvpiqheksnikvfihel 121  
 OY 146 NAGNOLSHMSPIYTIEMGDELLAKLRDAPFVRAHESNEMOPTLAISHAGSVVMAQTQ 205  
 |||||||  
 DB 122 naqngslmspiyltiemgdellaklardatffvrahesnempplalishagsvvmaq 181  
 OY 206 PREKRMESEWASGKVLCLDPLDGVYNYLAQORCNLDITWEGKIRYV 252  
 |||||||  
 DB 182 prtkrsewasgkvlclldpldgvynylaqrcnlditwegkiryv 228

RESULT 11  
 AAR32455  
 ID AAR32455 standard; Protein; 414 AA.  
 XX  
 AC AAR32455;  
 XX  
 DT 20-JUL-1993 (first entry)  
 XX  
 DE PE amino acids 2-414.  
 XX  
 KW PE; Pseudomonas exotoxin; Influenza A virus; M1 matrix protein;  
 KW T7 polymerase; fusion; hybrid; pVC-PEB7; pVC-PEM1-2.  
 XX  
 OS Synthetic.  
 XX  
 PN EP532090-A.  
 PD 17-MAR-1993.  
 XX  
 PF 02-SEP-1992; 92EP-0202660.  
 XX  
 PR 09-SEP-1991; 91US-0756249.  
 XX  
 PA (MERI ) MERCK & CO INC.  
 XX  
 PI Donnelly JJ, Friedman A, Howe LA, Liu MA, Marshall MS;  
 PI Montgomery DL, Oloff AI, Shi X, Ulmer J;  
 XX WPI: 1993-087107/11.  
 DR N-PSDB: AAQ37108.  
 XX

PT Bacterial toxin-antigen protein conjugates - to elicit cytotoxic  
 PT T-lymphocyte immune response, used for preventing viral  
 PT infections, e.g. by influenza virus, HIV and human  
 PT papilloma:Virus  
 XX  
 XX Disclosure; Page 33-35; 85pp; English.  
 XX  
 CC Example 5 describes the construction of pVC-PEB7.  
 CC A control plasmid was constructed which encodes a T7 polymerase  
 CC driven gene fusion consisting of PE amino acids 2 to 414 followed by  
 CC termination codons. pVC-PEM1-2 was digested with SacII and EcoRI to  
 CC remove the M1 sequence. The vector was gel purified and ligated to  
 CC an oligonucleotide that builds back PE codon no. 414 followed by  
 CC termination signals shown in AAQ37893. The resulting construction  
 CC was named pVC-PEB7 (AAQ37108).  
 CC  
 XX  
 SQ Sequence 414 AA:

Query Match 90.3%; Score 1195; DB 14; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-116;  
 Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 AEEAFDLMNECAKACVLDKDGVRSSRMVDPADITNGGVLYHSMVLEGNDALKLAI 85  
 |||||||  
 DB 2 aeafdlwneccakacvldkdyvrssrmvdpaditnggvlyhsmvlegndalkl 61  
 OY 86 DNALSTSDGLTIRLGGVEPNKPVRYSTYRQAGSLSMLVPIGHEKSNIKVFIHEL 145  
 |||||||  
 DB 62 dna1stsdgltrlrgvgnkpvrystyrqagswslmvlvpiqheksnikvfihel 121  
 OY 146 NAGNOLSHMSPIYTIEMGDELLAKLRDAPFVRAHESNEMOPTLAISHAGSVVMAQTQ 205  
 |||||||  
 DB 122 naqngslmspiyltiemgdellaklardatffvrahesnempplalishagsvvmaq 181  
 OY 206 PREKRMESEWASGKVLCLDPLDGVYNYLAQORCNLDITWEGKIRYV 252  
 |||||||  
 DB 182 prtkrsewasgkvlclldpldgvynylaqrcnlditwegkiryv 228

RESULT 12  
 AAR36806  
 ID AAR36806 standard; Protein; 426 AA.  
 XX  
 AC AAR36806;  
 XX  
 DT 25-AUG-1993 (first entry)  
 XX  
 DE PE domains I and II fused to Influenza A virus Ma.  
 XX  
 KW Vaccine; cytotoxic T lymphocyte; CTL; Influenza A virus;  
 KW matrix protein; Ma; Pseudomonas exotoxin; cell recognition domain;  
 KW Translocation domain; anti-viral agent; fusion construct.  
 XX  
 OS Chimeric Pseudomonas aeruginosa.  
 OS  
 FH Key  
 FT Region 2..414 Location/Qualifiers  
 FT Region /note= "PE domains I and II"  
 FT Region 415..426 /note= "amino acids 57-68 of Influenza A Virus  
 FT Matrix protein"  
 XX  
 PN EP541335-A.  
 PD 12-MAY-1993.  
 XX  
 PF 04-NOV-1992; 92EP-0310067.  
 XX  
 PR 08-NOV-1991; 91US-0792507.  
 XX  
 PA (MERI ) MERCK & CO INC.

```

XX  Donnelly JJ, Friedman A, Hawe LA, Liu MA, Marshall MS;
PI  Montgomery DL, Oliff AA, Shi X, Ulmer J;
XX  WPI: 1993-154266/19.
DR  N-PSDB; AAQ41714.
XX  Recombinant DNA encoding bacterial toxin-antigen conjugates - are
PT  useful as vaccines against viral infections, tumours and
PT  parasites
XX
PS  Example 4; Page 25; 81pp; English.
XX
CC  An MI gene fragment (encoding amino acids 57-68 of influenza A virus
CC  matrix protein) was subcloned into BS-PE, a plasmid constructed by
CC  inserting a 1.3kb NruI/SacII fragment of Pseudomonas exotoxin into
CC  the domain I and II coding regions of Pseudomonas exotoxin into
CC  pBLuescript II SK restricted with HincII and SacII. The PE-derived
CC  portion of the hybrid protein allows internalisation of the protein
CC  by an antigen-presenting cell. The hybrid protein is then processed
CC  and an antigenic segment (i.e. the Influenza A virus matrix protein)
CC  is presented on the cell surface where it elicits an immune response.
XX
SQ  Sequence 426 AA:
XX
Query Match 90.3%; Score 1195; DB 14; Length 426;
Best Local Similarity 100.0%; Pred. No. 2.8e-116;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 26 AEEAFDLNMECAKACVLDLKDGVRSSMSVDPADITNGGCVLHYSVMEGGNDALKIAI 85
DB 2 aeeafdlmecaacavldlkdgvrssmsvdpaditnggcvlhyssmveggndalkiai 61
XX
QY 86 DNALSTSDGLTIRLEGVEPNKPYRYSTRQARSGWSLNLVPIGHEKPSNIXYFIHEL 145
DB 62 dnalstsdgltirlegvepnkpyrystrgarsgwslnwlvpiqheksnlixvfihel 121
XX
QY 146 NAGNOLSHMSPIYTIEMGDELLAKIARDATFFVRAHESNEMOPTLAISHAGSVYMAQTO 205
DB 122 nagnqlshmspiytliegmdeallakiaratffvrahesnemoptlaishagsvymaqtg 181
XX
QY 206 PREKRKSEWASGKVLCLDPLDGVYNYLAQORCNLDLTWEGKIYRV 252
DB 182 prekrksewasgkvlclldpldgvynylaqrcnldltwegkiyrv 228
XX
RESULT 13
AAR32454
ID AAR32454 standard; Protein; 426 AA.
XX
AC AAR32454;
XX
DT 20-JUL-1993 (first entry)
XX
DE PE(2-414)-Ma(57-68) hybrid protein.
XX
KW PE: Pseudomonas exotoxin; influenza A virus; M1; matrix protein;
KW ompA; leader; signal; fusion; hybrid; BS-PEMA-1; PVC45DFR+T;
KW PVC-ompA-PEMA-1; PVC-PEMA-1; T7 promoter; RBS; ribosome binding site;
KW Initiation sequence; build-back.
XX
XX Synthetic.
XX
XX OS
XX PN EP532090-A.
XX PD 17-MAR-1993.
XX PF 02-SEP-1992; 92EP-0202660.
XX PR 09-SEP-1991; 91US-0756249.
XX PA (MERI ) MERCK & CO INC.

```

```

XX  Donnelly JJ, Friedman A, Hawe LA, Liu MA, Marshall MS;
PI  Montgomery DL, Oliff AL, Shi X, Ulmer J;
XX  WPI: 1993-087107/11.
DR  N-PSDB; AAQ37892.
XX  Bacterial toxin-antigen protein conjugates - to elicit cytotoxic
PT  T-lymphocyte immune response, used for preventing viral
PT  infections, e.g. by influenza virus, HIV and human
PT  papilloma virus
XX
PS  Disclosure; Page 30-31; 85pp; English.
XX
CC  Example 4 describes the subcloning of PEMA from BS-PEMA-1 into
CC  pVC45DFR+T. The PEMA insert (AAQ37890) was prepd. by restricting
CC  BS-PEMA-1 with SacI and removing the 3' overhang by treatment with
CC  T4 DNA polymerase, then restricting with EcoRI and gel purifying in
CC  pVC45DFR+T was restricted with EcoRI and the 5' overhang filled in
CC  with Klenow enzyme treatment. It was subsequently restricted with
CC  ApaI and gel purified. The vector and fragment were ligated together,
CC  and the resulting construction was named PVC-ompA-PEMA-1.
CC  The ompA signal sequence was removed from PVC-ompA-PEMA-1 by
CC  digestion with XbaI and HindIII. An oligonucleotide fragment contg.
CC  the T7 promoter, ribosome binding site, initiation sequence and a
CC  build-back of the 5' end of the PE coding region (AAQ37891) was
CC  ligated into the vector. The resulting plasmid construct was named
CC  PVC-PEMA-1 and encodes a T7 polymerase-driven gene fusion consisting
CC  of PE amino acids 2 through 414 joined to influenza M1 amino acids
CC  57 to 68 (Ma) (AAQ37892).
XX
SQ  Sequence 426 AA:
XX
Query Match 90.3%; Score 1195; DB 14; Length 426;
Best Local Similarity 100.0%; Pred. No. 2.8e-116;
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 26 AEEAFDLNMECAKACVLDLKDGVRSSMSVDPADITNGGCVLHYSVMEGGNDALKIAI 85
DB 2 aeeafdlmecaacavldlkdgvrssmsvdpaditnggcvlhyssmveggndalkiai 61
XX
QY 86 DNALSTSDGLTIRLEGVEPNKPYRYSTRQARSGWSLNLVPIGHEKPSNIXYFIHEL 145
DB 62 dnalstsdgltirlegvepnkpyrystrgarsgwslnwlvpiqheksnlixvfihel 121
XX
QY 146 NAGNOLSHMSPIYTIEMGDELLAKIARDATFFVRAHESNEMOPTLAISHAGSVYMAQTO 205
DB 122 nagnqlshmspiytliegmdeallakiaratffvrahesnemoptlaishagsvymaqtg 181
XX
QY 206 PREKRKSEWASGKVLCLDPLDGVYNYLAQORCNLDLTWEGKIYRV 252
DB 182 prekrksewasgkvlclldpldgvynylaqrcnldltwegkiyrv 228
XX
RESULT 14
AAR40102
ID AAR40102 standard; Protein; 613 AA.
XX
AC AAR40102;
XX
DT 27-JAN-1994 (first entry)
XX
DE Pseudomonas exotoxin for site-specific mutation with unpaired CYS.
XX
KW Pseudomonas exotoxin; PE; diphteria toxin; DT; immunotoxin;
KW target site; cytotoxin; unpaired cysteine; receptor; binding site;
KW monoclonal antibody; ligand; cell surface; mutation;
KW steric unpaired cysteine; s.u.c.
XX
XX Pseudomonas aeruginosa.
XX
XX OS
XX FH Key Location/Qualifiers

```



FT	Misc-difference	20	/note-	"unpaired cysteine residue may replace Lys"
FT	Misc-difference	25	/note-	"unpaired cysteine residue may replace Ser"
FT	Misc-difference	88	/note-	"unpaired cysteine residue may replace Ser"
FT	Misc-difference	96	/note-	"unpaired cysteine residue may replace Ser"
FT	Misc-difference	158	/note-	"unpaired cysteine residue may replace Ser"
FT	Misc-difference	182	/note-	"unpaired cysteine residue may replace Ser"
FT	Misc-difference	188	/note-	"unpaired cysteine residue may replace Arg"
FT	Misc-difference	192	/note-	"unpaired cysteine residue may replace Ser"
FT	Misc-difference	223	/note-	"unpaired cysteine residue may replace Ser"
FT	Misc-difference	245	/note-	"unpaired cysteine residue may replace Lys"
FT	Misc-difference	245	/note-	"unpaired cysteine residue may replace Ser"

XX	
PN	WO9315113-A.
XX	
XX	05-AUG-1993.
PD	
XX	15-JAN-1993;
PE	93WO-US00358.
XX	
PR	24-JAN-1992; 92US-0825396.
XX	
PA	(TANOX) TANOX BIOSYSTEMS INC.
XX	
PI	Chang TW;
XX	
WPI:	1993-258616/32.
XX	

XX Site-specifically mutated cytotoxin(s) with an unpaired cysteine  
PT - such that conjugation of a binding mol. to the Cys blocks  
PT receptor binding used as Immuno:toxins for highly specific  
XX targetting  
XX  
PS Claim 3; Page 20-23; 30pp; English.  
XX  
CC The new mutated toxin has an unpaired cysteine residue in  
CC or near the cytotoxin's receptor-binding site, and retains the  
CC same receptor-binding ability and cytotoxicity as the native  
CC cytotoxins provided they are not conjugated with a binding mol.  
CC The toxins are cross-linked through the free SH group of their  
CC unpaired cysteine residues to binding mols. (including monoclonal  
CC antibodies, fragments and other ligands) to form immunotoxins, and  
CC these immunotoxins do not bind to the cell surface receptors which  
CC are bound by the native cytotoxin. However, when the cross-linker  
CC is cleaved and the binding mol. is released, the cytotoxin regains  
CC its receptor-binding ability and its cytotoxicity.  
XX  
XX Sequence 613 AA:

Query Match	90.38;	Score 1195;	DB 14;	Length 613;
Best Local Similarity	100.0%;	Pred. No. 4,8e-116;		
Matches 221;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	26	AEEF.FD.LWNEC.KACAK.VLDI.KDGVSSR.SR.VDPAT.IADTNGC.VL.HYSWV.LEG.CND.LK.LAI
Db	1	aeef.fdl.wne.cak.c.vld.i.kdgv.ssr.sr.vdp.at.iad.tng.c.vl.hys.wv.le.g.cnd.l.k.lai
Qy	86	DNAL.SITS.DGL.TIR.LEG.GVE.PNK.PV.RSY.SY.TRQ.ANG.SW.SL.ML.VP.IGHEK.PS.NIK.VF.THEL
Db	61	dnal.sts.dgl.tr.le.g.gv.e.pnk.pv.rsy.sy.trq.ang.sw.sl.ml.vp.ighe.kps.nik.vf.the.l
Qy	146	NAG.NQ.LS.MS.SY.TT.EW.GDE.LLAK.LAR.DAT.FF.VA.HES.NE.MO.PLAI.SHAG.VS.VYM.AOTQ
Db	121	na.gnq.l.s.ms.sy.tt.e.w.g.de.ll.ak.la.r.dat.ff.va.hes.ne.mo.plai.shag.v.s.vym.a.ot.q

[illegible]

XX	
XX	WO9315113-A.
PN	
XX	
PD	05-AUG-1993.
XX	
PF	15-JAN-1993;
XX	
PR	24-JAN-1992;
XX	
PA	(TANO-) TANOX BIOSYSTEMS INC.
XX	
PI	Chang TW;
XX	
DR	WPI; 1993-258616/32.

PT Site-specifically mutated cytotoxin(s) with an unpaired cysteine  
 PR - such that conjugation of a binding mol. to the Cys blocks  
 PR receptor binding used as immuno:toxins for highly specific  
 PR targeting  
 XX  
 XX  
 PS Claim 3; Page 20-23; 30pp; English.  
 XX  
 CC The new mutated toxin has an unpaired cysteine residue in  
 CC or near the cytotoxin's receptor-binding site, and retains the  
 CC same receptor-binding ability and cytotoxicity as the native  
 CC cytotoxin provided they are not conjugated with a binding mol.  
 CC The toxins are cross-linked through the free SH group of their  
 CC unpaired cysteine residues to binding mols. (including monoclonal  
 CC antibodies, fragments and other ligands) to form immunotoxins, and  
 CC these immunotoxins do not bind to the cell surface receptors which  
 CC are bound by the native cytotoxin. However, when the cross-linker  
 CC is cleaved and the binding mol. is released, the cytotoxin regains  
 CC its receptor-binding ability and its cytotoxicity.  
 XX  
 SQ Sequence 613 AA:

Query Match 90.3%; Score 1195; DB 14; Length 613;  
Best Local Similarity 100.0%; Pred. No. 4.8e-116;  
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0y	26	AEEADPLNNEKACAVLDLKDGRSSRMSVDPAIADTNOQGLAHTSMVLEGGNDALKLAI	85
Dd	1	aeedfdlmeckacavldldkgvrssrmvdpaiddtngvylhysmvlegndalklali	60
0y	86	DNALITSDGLITRLIEGGVEPNKPVRYSTRQARSSWMLNMLVPIGHEKSPNIKAFIHL	145
Dd	61	dnaalitedgllitrlieggvepnkpvrystrlqarsswmlnmlvpiigheknplikfihel	120

OY 146 NAGNOLSHMSPIYTIEMGDELLAKIARDATFFVRAHESNEMQPTLAIASHAGSVVMAQTQ 205  
|||  
Db 121 nagnqlshmspiytleemgdellakiaratffvraheshesnemqptlaishagsvvmaqtq 180  
OY 206 PRERKRWSEWASGKVLCLLDPLDGVYNTLAOORCNLDPTWEGKIYRV 252  
|||  
Db 181 prekrwsewasgkvlclldpldgvyntlaqrcnlddtwegkiyrv 227

Search completed: November 13, 2001, 14:27:43  
Job time: 89 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:26:50 ; Search time 26.87 seconds  
(without alignments)  
211.047 Million cell updates/sec

Title: US-09-412-558-3

Perfect score: 1323

Sequence: 1 MHLPWIPVASTGLAGG.....YLAQRCNLDDEGKIVRV 252

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTCDS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1223	100.0	638	US-09-047-148-2	Sequence 2, Appl1
2	1195	90.3	613	US-08-405-615-1	Sequence 1, Appl1
3	1195	90.3	613	US-08-461-234-1	Sequence 1, Appl1
4	1195	90.3	613	US-08-463-480-1	Sequence 1, Appl1
5	1195	90.3	614	US-08-225-224-1	Sequence 1, Appl1
6	1195	90.3	614	US-08-722-258-1	Sequence 1, Appl1
7	1195	90.3	614	PCT-US95-04468-1	Sequence 1, Appl1
8	1181	89.3	635	US-09-046-992-2	Sequence 2, Appl1
9	93.5	7.1	1684	US-08-665-259-25	Sequence 25, Appl1
10	93.5	7.1	1684	US-08-762-500-25	Sequence 25, Appl1
11	93.5	7.1	1704	US-08-762-500-75	Sequence 75, Appl1
12	88	6.7	380	US-08-609-572-4	Sequence 4, Appl1
13	88	6.7	380	US-08-841-751-4	Sequence 4, Appl1
14	88	6.7	380	US-08-846-340-4	Sequence 4, Appl1
15	88	6.7	380	US-08-846-344-4	Sequence 4, Appl1
16	86.5	6.5	566	US-08-920-095-3	Sequence 3, Appl1
17	86.5	6.5	566	PCT-US96-05800-3	Sequence 3, Appl1
18	86.5	6.5	569	US-08-467-822-26	Sequence 26, Appl1
19	86.5	6.5	569	US-08-432-697-26	Sequence 26, Appl1
20	86.5	6.5	569	US-08-466-248-26	Sequence 26, Appl1
21	86.5	6.5	607	US-08-752-307B-12	Sequence 12, Appl1
22	84.5	6.4	3729	US-08-804-227C-4	Sequence 4, Appl1
23	81.5	6.2	281	US-08-284-465-6	Sequence 28, Appl1
24	77.5	5.9	466	US-08-726-136-28	Sequence 28, Appl1
25	77.5	5.9	466	US-09-103-434-28	Sequence 28, Appl1
26	77.5	5.9	466	US-09-687-594-28	Sequence 28, Appl1
27	76.5	5.8	349	US-08-806-597A-14	Sequence 14, Appl1

28	76.5	5.8	349	US-08-970-428A-14	Sequence 14, Appl1
29	76.5	5.8	599	US-09-000-145-2	Sequence 2, Appl1
30	76	5.7	1489	US-08-417-210A-137	Sequence 137, App
31	76	5.7	1794	US-08-417-210A-140	Sequence 140, App
32	74.5	5.6	550	US-08-417-210A-137	Sequence 137, App
33	74.5	5.6	551	US-08-417-210A-143	Sequence 143, App
34	74.5	5.6	2647	US-08-583-562B-8	Sequence 8, Appl1
35	74.5	5.6	2647	US-08-779-113-8	Sequence 8, Appl1
36	74.5	5.6	337	US-08-871-572B-1	Sequence 1, Appl1
37	74	5.6	337	US-08-871-572B-4	Sequence 4, Appl1
38	74	5.6	337	US-09-066-046-27	Sequence 27, Appl1
39	74	5.6	435	US-09-066-047-13	Sequence 13, Appl1
40	74	5.6	435	US-08-588-163-2	Sequence 2, Appl1
41	73.5	5.6	224	US-09-111-070-2	Sequence 2, Appl1
42	73.5	5.6	224	US-08-849-764C-2	Sequence 2, Appl1
43	73.5	5.6	224	PCT-US94-14498A-2	Sequence 2, Appl1
44	73.5	5.6	352	US-07-923-260A-6	Sequence 6, Appl1
45	73.5	5.6	352	US-07-923-260A-6	Sequence 6, Appl1

## ALIGNMENTS

RESULT 1  
US-09-047-148-2  
; Sequence 2, Application US/09047148  
; Patent No. 6086900  
; GENERAL INFORMATION:  
; APPLICANT: Draper, Rockford  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR USING  
; TITLE OF INVENTION: MEMBRANE-PENETRATING PROTEINS TO CARRY MATERIALS ACROSS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: United States  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/047,148  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION:  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 60/042,056  
; FILING DATE: 26-MAR-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Highlander, Steven L.  
; REGISTRATION NUMBER: 37,642  
; REFERENCE/DOCKET NUMBER: UTSP:072  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 638 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-09-047-148-2

Query Match 100.0%; Score 1323; DB 3; Length 638;  
Best Local Similarity 100.0%; Pred No. 3e-137;  
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MHLPWIPVASTGLAGSSASAEAFDLWNECAKACVLDLKDGVRSRMSVPAIA 60

Db 1 MHLPWIPVLASIGLGLAGSSASAEAFDLMNECAKACVLDKDGVRSSRMVDPALAI 60  
QY 61 DTNGGVVHYSMVLEGGNDALKLAIIDNALSTSDGLTRLEGVEPNKPVRYSTYRQARG 120  
Db 61 DTNGGVVHYSMVLEGGNDALKLAIIDNALSTSDGLTRLEGVEPNKPVRYSTYRQARG 120  
QY 121 SMSLMLVPIGHEKPSNIKVFIELHAGNOLSHMSPIYTIEMGDELLAKLARDAFFVRA 180  
Db 121 SMSLMLVPIGHEKPSNIKVFIELHAGNOLSHMSPIYTIEMGDELLAKLARDAFFVRA 180  
QY 181 HESNEMOPLAISAGVSVMAQOTPRREKRSEWASGKVCCLDPLDGVNYNYLAOQRN 240  
Db 181 HESNEMOPLAISAGVSVMAQOTPRREKRSEWASGKVCCLDPLDGVNYNYLAOQRN 240  
QY 241 LDDTWEKIRYV 252  
Db 241 LDDTWEKIRYV 252

## RESULT 2

US-08-405-615-1  
; Sequence 1, Application US/08405615  
; Patent No. 5602095  
; GENERAL INFORMATION:  
; APPLICANT: Pastan, Ira  
; APPLICANT: Fitzgerald, David J.  
; TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with  
; TITLE OF INVENTION: Increased Activity  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ellen L. Weber  
; STREET: One Market Plaza, Steuart Tower, Suite 2000  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,615  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/901,709  
; ATTORNEY/AGENT INFORMATION:  
; FILING DATE:  
; NAME: Weber, Ellen L.  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 15280-36  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-543-9600  
; TELEFAX: 415-543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 613 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-405-615-1

Query Match 90.3%; Score 1195; DB 1; Length 613;  
Best Local Similarity 100.0%; Pred. No. 3.8e-123; Indels 0; Gaps 0;  
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AEEAFDLMNECAKACVLDKDGVRSSRMVDPALADTNGGVVHYSMVLEGGNDALKLAI 85  
Db 1 AEEAFDLMNECAKACVLDKDGVRSSRMVDPALADTNGGVVHYSMVLEGGNDALKLAI 60

QY 86 DNALSTSDGLTRLEGVEPNKPVRYSTYRQARGSMSLMLVPIGHEKPSNIKVFIELH 145  
Db 61 DNALSTSDGLTRLEGVEPNKPVRYSTYRQARGSMSLMLVPIGHEKPSNIKVFIELH 120  
QY 146 NAGNOLSHMSPIYTIEMGDELLAKLARDAFFVRAHESNEMOPLAISAGVSVMAQOT 205  
Db 121 NAGNOLSHMSPIYTIEMGDELLAKLARDAFFVRAHESNEMOPLAISAGVSVMAQOT 180  
QY 206 PRREKRSEWASGKVCCLDPLDGVNYNYLAOQRNLDDTWEKIRYV 252  
Db 181 PRREKRSEWASGKVCCLDPLDGVNYNYLAOQRNLDDTWEKIRYV 227

## RESULT 3

US-08-461-234-1  
; Sequence 1, Application US/08461234  
; Patent No. 5821238  
; GENERAL INFORMATION:  
; APPLICANT: Pastan, Ira H.  
; APPLICANT: Fitzgerald, David J.  
; TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with  
; TITLE OF INVENTION: Increased Activity  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Crew  
; STREET: One Market Plaza, Steuart Street Tower  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105-1492  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/461,234  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/901,709  
; FILING DATE: 18-JUN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,615  
; FILING DATE: 15-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen Lauver  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 15280-36-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 543-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 613 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; US-08-461-234-1

Query Match 90.3%; Score 1195; DB 2; Length 613;  
Best Local Similarity 100.0%; Pred. No. 3.8e-123; Indels 0; Gaps 0;  
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AEEAFDLMNECAKACVLDKDGVRSSRMVDPALADTNGGVVHYSMVLEGGNDALKLAI 85  
Db 1 AEEAFDLMNECAKACVLDKDGVRSSRMVDPALADTNGGVVHYSMVLEGGNDALKLAI 60  
QY 86 DNALSTSDGLTRLEGVEPNKPVRYSTYRQARGSMSLMLVPIGHEKPSNIKVFIELH 145

DB 61 DNALSTSDGLTIRLEGGVPPNKPVRYSYTRQARSGMSLNLVLPICHEKSNIKVFTHL 120  
QY 146 MAGNOLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMOPTLAISHAGSVYMAQTQ 205  
DB 121 MAGNOLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMOPTLAISHAGSVYMAQTQ 180  
QY 206 PRERKRWSEWASGKVCLLDPLDGVYNYLAQRCNLDPTWEGKIYRV 252  
DB 181 PRERKRWSEWASGKVCLLDPLDGVYNYLAQRCNLDPTWEGKIYRV 227

## RESULT 4

US-08-463-480-1  
; Sequence 1, Application US/08463480  
; Patent No. 5854044  
; GENERAL INFORMATION:  
; APPLICANT: Pastan, Ira H.  
; APPLICANT: Filzgerald, David J.  
; TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with  
; TITLE OF INVENTION: Increased Activity  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,480  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/901,709  
; FILING DATE: 18-JUN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,615  
; FILING DATE: 15-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen Lauver  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 15280-36-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0300  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 613 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
US-08-463-480-1

Query Match 90.3%; Score 1195; DB 2; Length 613;  
Best Local Similarity 100.0%; Pred. No. 3.8e-123;  
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AAEAFDLMNECAKACVLDLKDGVSSRMSVDPALADTNGOGVLHYSVWLEGNDALKLAI 85  
DB 1 AAEAFDLMNECAKACVLDLKDGVSSRMSVDPALADTNGOGVLHYSVWLEGNDALKLAI 60  
QY 86 DNALSTSDGLTIRLEGGVPPNKPVRYSYTRQARSGMSLNLVLPICHEKSNIKVFTHL 145  
DB 61 DNALSTSDGLTIRLEGGVPPNKPVRYSYTRQARSGMSLNLVLPICHEKSNIKVFTHL 120

QY 146 MAGNOLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMOPTLAISHAGSVYMAQTQ 205  
DB 121 MAGNOLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMOPTLAISHAGSVYMAQTQ 180  
QY 206 PRERKRWSEWASGKVCLLDPLDGVYNYLAQRCNLDPTWEGKIYRV 252  
DB 181 PRERKRWSEWASGKVCLLDPLDGVYNYLAQRCNLDPTWEGKIYRV 227

## RESULT 5

US-08-225-224-1  
; Sequence 1, Application US/08225224  
; Patent No. 5635599  
; GENERAL INFORMATION:  
; APPLICANT: PASTAN, Ira  
; APPLICANT: KREITMAN, Robert J.  
; TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND  
; TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS  
; NUMBER OF SEQUENCES: 57  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend Kourie and Crew  
; STREET: Stewart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94105-1493  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/225,224  
; FILING DATE: 8-APR-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weber, Ellen L.  
; REGISTRATION NUMBER: 32,762  
; REFERENCE/DOCKET NUMBER: 15280-193  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 543-9600  
; TELEFAX: (415) 543-5043  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 614 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..614  
; OTHER INFORMATION: /label= native-pe  
US-08-225-224-1

Query Match 90.3%; Score 1195; DB 1; Length 614;  
Best Local Similarity 100.0%; Pred. No. 3.8e-123;  
Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 AAEAFDLMNECAKACVLDLKDGVSSRMSVDPALADTNGOGVLHYSVWLEGNDALKLAI 85  
DB 2 AAEAFDLMNECAKACVLDLKDGVSSRMSVDPALADTNGOGVLHYSVWLEGNDALKLAI 61  
QY 86 DNALSTSDGLTIRLEGGVPPNKPVRYSYTRQARSGMSLNLVLPICHEKSNIKVFTHL 145  
DB 62 DNALSTSDGLTIRLEGGVPPNKPVRYSYTRQARSGMSLNLVLPICHEKSNIKVFTHL 121  
QY 146 MAGNOLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMOPTLAISHAGSVYMAQTQ 205  
DB 122 MAGNOLSHMSPIYTIEMGDELLAKLARDATFFVRAHESNEMOPTLAISHAGSVYMAQTQ 181

QY 206 PREKRSEWASGKVLCLDPLDGVYNYNLAQQRCLDDTWEGKIYRV 252  
 |||||  
 182 PREKRSEWASGKVLCLDPLDGVYNYNLAQQRCLDDTWEGKIYRV 228  
 Db

## RESULT 6

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	90.3%;	Score 1195;	DB 3;	Length 614;
Best Local Similarity	100.0%;	Pred. No. 3.8e-123;		
Matches 227;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 86 DNALSTSDGTLIRLEGVEPNKPVRYSYTRQANGSNLMLVPDIGHKESNKKVFTHL 145  
|||||  
Db 62 DNALSTSDGTLIRLEGVEPNKPVRYSYTRQANGSNLMLVPDIGHKESNKKVFTHL 1212  
QY 146 NAGNQLDSMSPTYYTTEMGDELLAKLARDATFTFYVAHESNEMQPTLASHAGVSVVMAQTQ 205

Db	Qy	Db
122	PREKRMSWASGKYLCLDPLDGYNYNTLAQQRNLDPTWEGKIRYV 252	122
182	PREKRMSWASGKYLCLDPLDGYNYNTLAQQRNLDPTWEGKIRYV 228	182

## RESULT 7

```

PCT-US95-04468-1
: Sequence 1, Application PC/TUS9504468
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND
: TITLE OF INVENTION: CIRCULARLY PERMUTED FUSION PROTEINS
: NUMBER OF SEQUENCES: 59
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/04468
: FILING DATE: 07-APR-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/225,224
: FILING DATE: 08-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Weber, Ellen L.
: REGISTRATION NUMBER: 32,762
: REFERENCE/DOCKET NUMBER: 15280-193-1PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 614 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..614
: OTHER INFORMATION: /label= native-PE
PCT-US95-04468-1

```

Query Match	90.3%;	Score 1195;	DB 5;	Length 614;
Best Local Similarity	100.0%;	Pred. NO. 3.8e-123;		
Matches 227;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 146 MAGNOLSHSPIYTIEMGDELLAKLRDAPFFVVRHESHENQPIILASHAGVSVMQ 205  
 DB 62 DNALSTTSGGLTRLEGGEVPEKPKPVRYSTTRQARGSMGLNMLVPIGHEKPSNKIVFTHEL 121  
 QY 86 DNALSTTSGGLTRLEGGEVPEKPKPVRYSTTRQARGSMGLNMLVPIGHEKPSNLVFTHEL 146  
 DB 122 MAGNOLSHSPIYTIEMGDELLAKLRDAPFFVVRHESHENQPIILASHAGVSVMQ 181

QY 206 PREKRWSEWASGKVLCLDLPIDGVNYYLAQQRCNDDTWEGKIYRV 252  
|||||  
Db 182 PREKRWSEWASGKVLCLDLPIDGVNYYLAQQRCNDDTWEGKIYRV 228

## RESULT 8

US-09-046-992-2  
; Sequence 2, Application US/09046992

```
; Patent No. 6140066
; GENERAL INFORMATION:
; APPLICANT: Lorberbom-Galekl, Haya
; APPLICANT: Yarkoni, Shai
; APPLICANT: Ben-Yehudah, Ahmi
; TITLE OF INVENTION: METHODS OF CANCER DIAGNOSIS
; TITLE OF INVENTION: USING A CHIMERIC TOXIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,992
; FILING DATE: 24-MAR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 9457-0013-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 635 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-046-992-2
```

```
Query Match      89.3%; Score 1181; DB 4; Length 635;
Best Local Similarity 95.3%; Pred. No. 1.4e-121;
Matches 223; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 19 GSSASAAEAPDLMNECAKACVLDKDCVRSRMSVDPALADTNGCVLHYSMTVEGNN 78
DB 16 GSGGGGSGQAPDLNMECAKACVLDKDCVRSRMSVDPALADTNGCVLHYSMTVEGNN 75
QY 79 DALKLAIIDNALITSIDGLTIRLEGVEPNKPVRYSYTRQARGSWSLNMVLPICHEKPSNI 138
DB 76 DALELAIDNALITSIDGLTIRLEGVEPNKPRLRYSTRQARGSWSLNMVLPICHEKPSNI 135
QY 139 KYFIHELNAGNOLSHMSPTTYTTEMDELLAKLARATFFVRAHESNEMOPTLAISHAGVS 198
DB 136 KYFIHELNAGNOLSHMSPTTYTTEMDELLAKLARATFFVRAHESNEMOPTLAISHAGVS 195
QY 199 VYMAOTOPRREKRMSWASGVKVLCLDPLDGYVNYLAQRCLNDPDWEKIRYV 252
DB 196 VYMAOTOPRREKRMSWASGVKVLCLDPLDGYVNYLAQRCLNDPDWEKIRYV 249
```

```
RESULT 9
US-08-665-259-25
; Sequence 25, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
```

```
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: 165-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1684 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-665-259-25
```

```
Query Match      7.1%; Score 93.5; DB 3; Length 1684;
Best Local Similarity 22.9%; Pred. No. 0.59;
Matches 36; Conservative 22; Mismatches 54; Indels 45; Gaps 4;

QY 114 YTRQARGSNLSLNVLPICHEKPSNIKVFTH-----ELANGNOLSHMSPTTYTEMGDE 165
DB 729 FLKQYAGAGYHMTLVKEPCNPEDISQVLNHNHPNATLSSAGAEISFTLPRESTHREFG 788
QY 166 LIAKLARATFFVRAHESNEMOPTLAISHAGVSV-----MA 202
DB 789 LRAKL-----EKKQKELGAIASFGASITTMEEVFLRVGKLVDSMDIOAIQLP 835
QY 203 OTOPRREKRMSWASGVKVLCLDPLDGYVNYLAQR 238
DB 836 ALQYQERRASDPAVDSNLCGAMDPSDIGALIEEER 872
```

```
RESULT 10
US-08-762-500-25
; Sequence 25, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
```







Best Local Similarity	20.76	22.02	22.02
Matches	34	Conservative	26
		Mismatches	46
		Indels	58
		Gaps	7

Search completed: November 13, 2001, 14:26:52  
Job time: 38 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:28:19 : Search time 30.29 Seconds  
(without alignments)  
633.740 Million cell updates/sec

Title: US-09-412-558-3

Perfect score: 1323  
Sequence: 1 MHLIPHWIPLVASIGLLAGG.....YLAQRCLDDPTWEGKIYRV 252

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1323	100.0	638	2 A30347	exotoxin A precurs
2	1307	98.8	638	2 C83503	exotoxin A precurs
3	93.5	7.1	1704	2 A59188	ATP-binding casset
4	93.5	7.1	1704	2 S71363	probable ATP-bind
5	93.5	7.1	4180	2 G83559	hypothetical prote
6	90.5	6.8	311	2 S43799	hypothetical prote
7	90	6.8	531	2 S20900	titin - mouse (frr
8	89.5	6.8	521	2 S40933	probable replicati
9	89	6.7	380	2 S62525	hypothetical prote
10	89	6.7	6805	2 S20901	titin - rabbit (fr
11	88	6.7	26926	1 I38344	titin, cardiac mus
12	88	6.7	915	2 T21147	hypothetical prote
13	88	6.7	1195	2 T43735	pyruvate carboxyla
14	87	6.6	1064	2 F86182	hypothetical prote
15	87	6.6	1444	2 S57335	cleavage and polya
16	86.5	6.5	569	1 URCBP	urease (EC 3.5.1.5
17	86.5	6.5	1040	2 A49356	transient axonal g
18	85.5	6.5	408	2 H86891	exported serine pr
19	85.5	6.5	695	2 S44049	nucleocapsid prote
20	85.5	6.5	2035	2 T30259	multiple PDZ domai
21	85	6.4	441	2 S74626	erythroid ankyrin
22	84.5	6.4	559	2 F71420	hypothetical prote
23	84.5	6.4	560	2 T48841	heat shock protein
24	84.5	6.4	695	1 VHTMV	nucleocapsid prote
25	84	6.3	984	2 T44608	pyruvate carboxyla
26	83.5	6.3	442	1 E69997	nitrolicarboxylate
27	83	6.3	1150	2 A83978	pyruvate carboxyla
28	83	6.3	1588	2 B85547	probable RTX famil
29	83	6.3	9376	2 T14593	stringomyein synth

30	82.5	6.2	344	2 F83620	probable binding p
31	82	6.2	325	2 B69786	thiamin-monophosph
32	82	6.2	338	2 T36025	conserved hypothet
33	82	6.2	1807	2 JC6319	integrin beta-4 ch
34	81.5	6.2	274	2 E75303	citrate lyase, bet
35	81.5	6.2	281	1 PMSAP	beta-lactamase (EC
36	81	6.1	401	2 S50533	hypothetical prote
37	81	6.1	788	1 S15239	anthranilate synth
38	81	6.1	986	1 OYURGA	spectat receptor p
39	81	6.1	997	2 T32814	hypothetical prote
40	81	6.1	1295	2 T24587	hypothetical prote
41	80.5	6.1	1034	2 S60051	sodium channel alp
42	80.5	6.1	1034	2 S60060	sodium channel alp
43	80.5	6.1	1700	2 G86131	probable invasin 2
44	80	6.0	502	2 T26851	hypothetical prote
45	80	6.0	998	2 T14764	hypothetical prote

## ALIGNMENTS

```
RESULT 1
A30347
exotoxin A precursor - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 24-Nov-1999
C:Accession: A30347
R:Gray, G.L.; Smith, D.H.; Balderidge, J.S.; Harkins, R.N.; Vasil, M.L.; Chen, E.Y.; H
Proc. Natl. Acad. Sci. U.S.A. 81, 2645-2649, 1984
A>Title: Cloning, nucleotide sequence, and expression in Escherichia coli of the exot
A:Reference number: A30347, MUID:84194053
A:Accession: A30347
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-638 <GRA>
A:Cross-references: GB:K01397; GB:M23348; NID:9151215; PIDN:AAB59097.1; PID:9151216
C:Superfamily: Pseudomonas aeruginosa exotoxin A
C:Keywords: exotoxin

Query Match 100.0%; Score 1323; DB 2; Length 638;
Best Local Similarity 100.0%; Pred. No. 9.7e-109;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MHLIPHWIPLVASIGLLAGSSASAAEFDLMECKACVLDLKQGVRSRMSVDAIA 60
    |||||||
DB 1 MHLIPHWIPLVASIGLLAGSSASAAEFDLMECKACVLDLKQGVRSRMSVDAIA 60

OY 61 DTNCGVLHSMVLEGGNDALKLAIDNALSTSDGLFTRLEGGVEPNKPVRYSYTRQARG 120
    |||||||
DB 61 DTNCGVLHSMVLEGGNDALKLAIDNALSTSDGLFTRLEGGVEPNKPVRYSYTRQARG 120

OY 121 SWSLNLVLPITGHEKPSNIKVFIFHELNAGNOLSHMSPIYTTMEGDELLAKLRADTFEVR 180
    |||||||
DB 121 SWSLNLVLPITGHEKPSNIKVFIFHELNAGNOLSHMSPIYTTMEGDELLAKLRADTFEVR 180

OY 181 HESNEMOPTLAISHAGSVYMAOTQPRRKRMSMASGKVLCLDPLDGVNYNTAQRGN 240
    |||||||
DB 181 HESNEMOPTLAISHAGSVYMAOTQPRRKRMSMASGKVLCLDPLDGVNYNTAQRGN 240

OY 241 LDDTWEGKIYRV 252
    |||||||
DB 241 LDDTWEGKIYRV 252

RESULT 2
C83503
exotoxin A precursor PA1148 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83503
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Watterner, P.; Hickey, M.J.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
```

.: Lory, S.: Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: C83503  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-638 <SNO>  
 A:Cross-references: GB:AE004544; GB:AE004091; NTD:g9947060; PIDN:AAC04537.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: toxA; PA1148  
 C:Superfamily: *Pseudomonas aeruginosa* exotoxin A

Query Match	98.8%;	Score 1307;	DB 2;	Length 638;
Best Local Similarity	98.8%;	Pred. No. 2.5e-107;		
Matches 249; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

Oy 1 MHLIPHWIPLVASLGLLAGGSSASNAAEAPDLWNECKACVLDKDGVRSSRMSYDPAIA 60  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 1 MHLTPHWIPLVASLGLLAGGSSAFSAFAEEAPDLWNECKACVLDKDGVRSSRMSYDPAIA 60

Qy 61 DTNGQGVLAHSMVLEGGNDALKIADNLSTISDGLTRLEGGVEPKPRVSYTQANG 12  
 |||||  
 Db 61 DTNGQGVLAHSMVLEGGNDALKIADNLSTISDGLTRLEGGVEPKPRVSYTQANG 12

[illegible]

Accession	Protein	Length
W1	102 HESNMQPTLAISAGSYVMAQAQPRREK	102
Db	181 HESNMQPTLAISAGSYVMAQAQPRREK	181
07	241 LDDTWEGSKTYRV	252

Db 241 LDDTWECKITRV 252

ATP-binding cassette transporter ABC3 - human  
A59188  
C1:Species: Homo sapiens (man)  
C2:Date: 18-Feb-2000 #sequence revision 18-Feb-2000 #text change 04-Mar-2000

C/Accession: A59188  
R/Comments: T.D.; Van Raay, T.J.; Petry, L.R.; Klingler, K.W.; Landes, G.M.  
Genomics 39, 231-234, 1997  
Ar/Title: The cloning of a human ABC gene (ABC3) mapping to chromosome 16p

A: Accession: A59188  
A: Reference number: A59188; MOLDB-5117225  
A: Status: preliminary; not compared with conceptual translation  
A: Molecule type: mRNA  
A: Accession: 141704; CON5

A:Cross-references: GB:U78735; NID:g1699037; PIDN:AAC50967.1; PID:g1699037  
C:Genetics:  
A:Gene: GDB:ABC3  
A:Cross-references: GDB:3770735; OMIM:601615

A:Map position: 10p13.3-10p13.3  
C:Superfamily: ATP-binding cassette homology

114 YTRQARGSWLNWLVLPICHEKPSNIKYFIH-----EINAGNOLSHNSPIYTIEWGDE 1

```

Db      749  FLTKQYGGAGYNNHTLYKPEHCNPEIDSLYNNHNPNATLSSAGAEELSTLPRESIHREG 8
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy     166  LLAKLARDAFFVVRKHNSENENQPTLASHAGSYV-----MA 2

```

```
Db      809 -----EKKQKELGIASFGASITTWEEVFLRVGKLDDSDMDIOAIOLP 8
```

QY 203 QTPRRREKRWEEMASGKVLCTLDPLDGVNYLAQR 238  
 : : : : : : : : : :  
Db 856 ALQYHERRASDAVDSNLTCAMDPSDIGALIEER 892

RESULT 4  
S71363  
probable ATP-binding cassette transporter ABC-3 - human

C:Species: Homo sapiens (man)  
C:Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 02-Feb-2001  
C:Accession: S71363

FEBS Lett. 391, 61-65, 1996  
A7:Title: Primary structure of a novel ABC transporter with a chromosomal localization  
A:Reference number: S71363; MUID: 96326608  
A:Accession: S71363

A: Molecule type: mRNA  
A: Residues: 1-1704 <KLU>  
A: Cross-references: EMBL: X97187; NID: g151454529; PIDN: CAA65825.1; PTD: e243436; PTD: g151

C;Genetics: A;Gene: GDB:ABC3  
A;Cross-references: GDB:3770735; OMIM:601615  
A;Map position: 16p13.3-16p13.3

C:Superfamily: transmembrane protein  
C:Keywords: ATP binding; nucleotide binding; p-loop; phosphoprotein; transmembrane protein  
E:455-483/Domain: transmembrane #status predicted <TM>  
F:307-329/Domain: transmembrane #status predicted <TM>

F: 313-334/Domain	transmembrane	#status	predicted	<TM>
F: 373-394/Domain	transmembrane	#status	predicted	<TM>
F: 401-422/Domain	transmembrane	#status	predicted	<TM>
F: 450-475/Domain	transmembrane	#status	predicted	<TM>
F: 540-730/Domain	ATP-binding cassette	homology	ABC1	

F	#685-690/Region:	nucleotide-binding motif B	<TM
F;1100-1120/Domain: <td>transmembrane</td> <td>#status predicted</td> <td>&lt;TM</td>	transmembrane	#status predicted	<TM
F;1145-1169/Domain: <td>transmembrane</td> <td>#status predicted</td> <td>&lt;TM</td>	transmembrane	#status predicted	<TM

F;1215-1236/Domai:	transmembrane#status predicted<ITM
F;1245-1264/Domai:	transmembrane#status predicted<ITM
F;1299-1324/Domai:	transmembrane#status predicted<ITM
P;1300-1324/Domai:	amp-binding cassette homology ABC

Protein	Accession	Region	Binding site	Covalent	CAMP-dependent	status
F.1116	1423	Region: nucleotide-binding motif B (loop)				
F.1535	1540	Region: nucleotide-binding motif B				
F.674	866	154	Binding site: phosphate (Ser)	(covalent)	(by CAMP-dependent kinase)	#
F.1344	Binding site: phosphate (Thr)	(covalent)	(by CAMP-dependent kinase)	status	pro	

Query Match	7.1%	Score 93.5;	DB 2;	Length 1704;
Best Local Similarity	22.9%	Pred. No. 12;		
Number of Mismatches	73	54;	Indels 45;	Gaps 4;

QY 114 YTRQARCSWSLNLPVIGCHEKPSNIKVFIH-----ELNAGNQLSHMSPIYTIEMGDE 165  
: : : : : : : : : : : : : : : :  
Db 749 FIKOKYAGYHMTLVKEPHCNPEDISQLVHHHVNPATLESAGAEISFILPRESTRHEEG 808

```

Qy      166 LIAKLARDATFEVRAHESNEMOPLAISAGVSV-----MA 202
          |||      : | | : | | :
Db      809 LEAKL-----EKKQKELGIAFGASITTMEEVFLRVGKLVDSMDIAIQLP 855

```

QY 203 QTQPRRERRWSEWASGKVLCLDLPDDGVINLQAQR 256  
+ : + : + : + : + : + : + :  
Db 856 ALQYQHERRASDWAYDSNLGCAMPSPDGICALIEER 892

RESULT 5  
G83559  
hypothetical protein PA0690 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C:\Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: G83559  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lartig, K.; Llm, .; Lofy, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: G83559  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-4180 <STC>  
 A:Cross-references: GB:AE004504; GB:AE004091; NID:99946568; PIDN:AAG04079.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA0690

Query Match 7.1%; Score 93.5; DB 2; Length 4180;  
 Best Local Similarity 27.8%; Pred. No. 40;  
 Matches 50; Conservative 22; Mismatches 81; Indels 27; Gaps 7;

OY 44 LKDVSSKRSVDPADTNGGVLHYSWLEGGNDALKAIDNALSITSDDG----- 95  
 Db 3257 LCGNLISMRPGDAGNIAPRGDSIPSGNLNPSQGLVLAAGTGRITSDGALQGGG 3316  
 OY 96 -LTIRLEGVEPKPVRYSTYRQARGSWSLNMLVPIGHEKPSNITKVFHEL-NAGNOLSH 153  
 Db 3317 DLNVRIGCEVNPISREARATQYSSSG---FDGLYSGG-----TINLQCALINLRG 3364  
 OY 154 MSPIYTIEMG--DELLAKLARDATFFVRAHESNEMOPTLAISHAGSVYVMAQTQPRREKR 211  
 Db 3365 SASLYSGALGIDPRVDTLRLDPA-EVRSRDA--FSPTLASTGCTLTIVAGDTGMRLER 3421

RESULT 6

S43799  
 hypothetical protein 34.1 - phage SPPI

C:Species: phage SPPI

C:Date: 25-Dec-1994 #sequence\_revision 26-May-1995 #text\_change 17-Nov-2000

C:Accession: S43799; T42331; S41173

R:Pedre, X.; Weise, F.; Chal, S.; Lueder, G.; Alonso, J.C.

J. Mol. Biol. 236, 1324-1340, 1994

A:Title: Analysis of cis and trans acting elements required for the initiation of DNA re

A:Reference number: S43798; MUID:94172631

A:Accession: S43799

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-311 <PEO>

A:Cross-references: EMBL:X67865; NID:9472886; PIDN:CAA48050.1; PID:9439630

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1992

R:Alonso, J.C.; Luder, G.; Steige, A.C.; Chal, S.; Weise, F.; Trautner, T.A.

Gene 204, 201-212, 1997

A:Title: The complete nucleotide sequence and functional organization of *Bacillus subtilis*

A:Reference number: Z22137; MUID:98094274

A:Accession: T42331

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-136, 'Y', 138-311 <ALO>

A:Cross-references: EMBL:X97918; PIDN:CAA65538.1

Query Match

Best Local Similarity 6.8%; Score 90.5; DB 2; Length 311;  
 Matches 35; Conservative 21; Mismatches 44; Indels 51; Gaps 6;

OY 52 RMSVDPAID-TNGGVLHYSWLEGGNDALKAIDNALSITSDDGLTIRLEGVEPKPV 110  
 Db 112 RTMLDGVLDHENGSGLE-----ITASESLKEEWEGEDIPMO-- 150  
 OY 111 RYVYTGARSGNSLMLVPIGHEKPSNITKVFHELMNAGNOLSHSPITYTMCMD 164  
 Db 151 -----YMIQVOH-----NIKVVEADFAVVALIGCNKYKH---YTERDD 187

OY 165 ELLAKLARDATFFVRAHESNEMOPTLAISHA 195  
 Db 188 ELISHYQGEHYHYFMNHNFSANIPPTASDSA 218

RESULT 7  
 S20900

titin - mouse (fragment)

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 18-Jun-1999

C:Accession: S20900

R:Labell, S.; Gautel, M.; Lakey, A.; Trinick, J.

EMBO J. 11, 1711-1716, 1992

A:Title: Towards a molecular understanding of titin.

A:Reference number: S20897; MUID:92258380

A:Accession: S20900

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-531 <LAB>

A:Cross-references: EMBL:X64700; NID:954807; PIDN:CAA45941.1; PID:954808

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992

C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology;

C:Keywords: muscle

Query Match 6.8%; Score 90; DB 2; Length 531;  
 Best Local Similarity 22.4%; Pred. No. 48;  
 Matches 50; Conservative 33; Mismatches 86; Indels 54; Gaps 10;

OY 56 DPAIADTNGGVLHYSWLEGGN--DALK--LAIDNALSITSDDGLTIRLEGVEPKPV 111  
 Db 159 DAAIIVDTSS---FTSLVLDNVRNYDSGKYTLTLENSGSKSAFVYRVLD--TPSPVN 213  
 OY 112 YSTYRQARGSWSLNMLVPIGHEKPSNITKVFHELMNAGNOLSHSPITYTMCDELLAKLA 171  
 Db 214 LKYTEITKDSVSTWEPPL-LDGSKIKYIVKREATRSYAAYVYCNKSNKIDQLQ 272  
 OY 172 RDATFFVRAHESNEM-----OP-----TLAIS-----HAG 196  
 Db 273 EGGSYFRVTAENEDYIGLAARTADPKVAEVPQPEKTVDDYTRNSVLSMTKPEHDC 352  
 OY 197 VSVV---MAQTQPRREKRWSEMAKVLCLLDPLDGYVNYLAQ 236  
 Db 333 GSKITQYIVEMQAKNTDKWSECARVK-----SLDAVITNLTQ 369

RESULT 8

S40933  
 probable replication licensing factor MCM6 - *Caenorhabditis elegans* (fragment)

N:Alternate names: hypothetical protein ZK632.1

C:Species: *Caenorhabditis elegans*

C:Date: 06-Jan-1995 #sequence\_revision 19-Jul-1996 #text\_change 24-Sep-1999

C:Accession: S40933

R:Barks, M.

submitted to the EMBL Data Library, February 1993

A:Reference number: S40933

A:Accession: S40933

A:Molecule type: DNA

A:Residues: 1-521 <BER>

A:Cross-references: EMBL:Z22181; NID:9297998; PID:9297999

A:Comment: The complex of six MCM proteins is one of several proteins that must be bound to phosphorylated and dissociate from the chromatin.

C:Genetics:

A:Map position: 3

A:Inserts: 463/2

C:Complex: The predominant form is a heterohexamer of MCM2, MCM3, MCM4, MCM5, MCM6, a

C:Function: part of the replication licensing system that permits DNA replication

C:Superfamily: cyclopropane-fatty-acyl-phospholipid synthase; MCM homology

C:Keywords: cell cycle control; DNA replication; methyltransferase; nucleus

F:1-365/Domain: MCM homology (fragment) <MCM>

	Query Match	6.8%	Score 89.5;	DB 2;	Length 521;	
	best Local Similarity	23.2%	Pred. No 5.2;	Mismatches 90;	Indels 53;	Gaps 13
	Matches 55;	Conservative 39;				
Qy	17 LAGGSASAAEEA--FDLMECA--KACVLDLKDVSRMSKVDPALADTNGO-----G	66				
Dd	19 LVGDAGCAVEEDYDILMSSKSTEDRALTKRMSDKIKENIVDSLEPNLYGHYEKLQ	78				
Qy	67 VLAHSVLLEGNDALKAIIDNALSTSDLTFLRLGG-----GVPEFNKRVRYSYT	115				
Dd	79 VL---LMHLLGG--VAKKSRDEGTSLMGD-INCVLGDPSTAKSQYLKAVEEFP-RAIYT	131				
Qy	116 ROARGSLSMLWLPICGEKPSTNIKVFTEHLNNGNSLHKSPITYITEMODELLAKIADAT	175				
Dd	132 -SGRASSAAGITAAVAYADES----FEFIEEGALLMDANGCCIDEEDKA-----DLK	180				
Qy	176 FEFVARSHSNNEOPTLATISHSVGAWAQAQPRRERKRWSEMAKGVCILLDDGVYN	232				
Dd	181 DQVALHEAMEGO-TISTTKAGVKATLIN-----ANASTLANAPNVNGRYD	223				

**RESULT**

9  
hypothetical protein SPAC8A.09c - fission yeast (*Schizosaccharomyces pombe*)  
S62525  
**C:**Species: Schizosaccharomycetes pombe  
**C:**Date: 16-May-1996 #sequence\_revision 13-Mar-1997 #text\_change 31-Jan-2000  
**C:**Accession: T39131; T41433; S62525  
R.iye, G.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, November 1995  
**A:**Reference number: Z21830  
**A:**Accession: T39131  
**A:**Status: preliminary; translated from GB/EMBL/DDBJ  
**A:**Molecule type: DNA  
**A:**Residues: 1-380 <LEVE>  
**A:**Cross-references: EMBL:Z66569; NID:g1052533; PIDN:CAB91519.1; PID:g1052540; SPDB:SPACC  
submitted to the EMBL Data Library, October 1998  
**A:**Reference number: Z21993  
**A:**Accession: T41433  
**A:**Status: preliminary; translated from GB/EMBL/DDBJ  
**A:**Molecule type: DNA  
**A:**Cross-references: EMBL:AL032824; PIDN:CAB37424.1; GSPDB:GN00066; SPDB:SPCC564.1lc  
**A:**Accession: F00000  
**A:**Cross-references: EMBL:F00000; NCBI:U00000; PIR:P00000  
**A:**Experimental source: strain 972h-, cosmid c584  
**C:**Genetics: <LYE1>  
**A:**Gene: SPAC8A.09c  
**A:**Introns: none  
**A:**Map position: 1  
**A:**Introns: 8/2  
**C:**Genetics: <SEEL>  
**A:**Gene: SPCC584.1lc  
**A:**Map position: 3

[illegible]

C:Species: *Oryctolagus cuniculus* (domestic rabbit)  
C:Date: 30-Sep-1993 #sequence\_revision 30-sep-1993 #text\_change 18-Jun-1999  
C:Accession: S20901; I46520  
R:Label1, S.; Gautel, M.; Lakey, A.; Trznick, J.  
EMBO J. 11, 1711-1716, 1992  
A:Title: Towards a molecular understanding of titin.  
A:Reference number: S20897; MUID:92258380  
A:Accession: S20901  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-6805 <LAB>  
A:Cross-references: EMBL:X64696  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992  
R:Label1, S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U.  
Nature 345, 273-276, 1990  
A:Title: A regular pattern of two types of 100-residue motif in the sequence of titin  
A:Reference number: I46520; MUID:90238553  
A:Accession: I46520  
A:Status: translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 4235-9250 <LAB>  
A:Cross-references: EMBL:X11329; NID:91756; PIDN:CAA35207.1; PID:9930251  
C:Superfamily: titin, fibronectin type III repeat homology; immunoglobulin homology;  
C:Keywords: muscle

Query Match	6.7%	Score 89;	DB 2;	Length 6805;
Best Local Similarity	22.1%	Pred. No. 1.9e+02;		
Matches 46;	Conservative 35;	Mismatches 79;	Indels 48;	Gaps

  

QY	56	DPAAIDTNGQGVLVNYSVLEGN-DALK-LAIDNALSTSDGTLTFLBEGVEPNKPYR	111
		-	-
DB	4847	DAALIDTSS--PTSLVDNVRYSOGKLTLELSSGTSKFAFVTVRLD--TPSPVN	4901
QY	112	YSYTPQANGSMLWLVLPDHEKPNRYKIVHELNQNGLSMSPSTIYTEMDELLAKLA	171
		-	-
DB	4902	LKVEITTDOSVSTIWEPP-LDGGSKIKNTIYKRRSTKSYAAVVTGNCHKSSMKIDQL	4960
QY	172	RDAPEFVRAHESNM-----QP-----PLAIS-----HAG	196
		:::	:::
DB	4961	ECCSYTFPVTAEENEYGIQTPARTADPIKVAEVPQPGKITVDVTNRSVSLWTKPEHDG	5020
QY	197	VSVV---MAOTQPPREKKWSMAGKVL	221
		-	-
DB	5021	GSKITQIYTEMQAKHSEKMSCAVYKSL	5048

RESULT 11  
 I38344  
 titin cardiac muscle [validated] - human  
 N:Alternate names: connectin  
 N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.-.)  
 C:Species: Homo sapiens (man)  
 C:Date: 12-Aug-1996 #sequence-revision 12-Aug-1996 #text-change 15-Sep-2000  
 C:Accession: I38344; I38345; S20898; S20897; S20899; S6365; S37393  
 R:Label: S.; Kolmeier, B.  
 Science 270, 293-296, 1995  
 A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.  
 A:Reference number: A57430; MUID:96026330  
 A:Accession: I38344  
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/  
 A:Molecule type: mRNA  
 A:Residues: 1-26926 <IABI>  
 A:Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425  
 R:Musco, G.; Tzatzalos, C.; Schuck, P.; Pastore, A.  
 Biochemistry 34, 553-561, 1995  
 A:Title: Dissecting titin into its structural motifs: Identification of an alpha-heli  
 A:Reference number: I38345; MUID:95119041  
 A:Accession: I38345  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1977-2014 <MUS>  
 A:Cross-references: EMBL:X83270; NID:g602579; PIDN:CA58243.1; PID:g602580





A:Molecule type: mRNA  
 A:Residues: 1-1195 <TUZ>  
 A:Cross-references: EMBL:L36530; PDB:AAB64306.1  
 C:Function:  
 A:Description: EC 6.4.1.1 [validated; MUID:97218699]  
 C:Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-binding  
 C:Keywords: ligase  
 F:1161/Binding site: biotin (Lys) (covalent) #status predicted

Query Match 6.7%; Score 88; DB 2; Length 1195;  
 Best Local Similarity 21.6%; Pred. No. 22; Indels 64; Gaps 11;  
 Matches 50; Conservative 34; Mismatches 84; Indels 64; Gaps 11;  
 QY 1 MHLPWIPVLVSLGLLAGSSASAAEEAF-----DLWNECAKA- 39  
 DB 697 LNLPLIL-----LGEAAGNNGVVEALISTGVSDPTKKRYDLKTYTNLADELIVKAG 751  
 QY 40 ---CVLDLKD-----GVRSKMSVDPAL-ADTNGGVLYHSWYLEGNDAL 81  
 DB 752 THLICIKDMAGLLKPOAKLLIAIREKHPDVIHITHDTSAGAGVASMALACAEKAGADVV 811  
 QY 82 KLAIDNALISTSD---GLTIR-----LEGVEPNKPVRYG-VTROARGSMGLNMLVPI- 130  
 DB 812 DVAVDMSGFTSOPSMGAVVASLQGTPLDITGLMLDISYSATWECTRLIYA-----PPE 866  
 QY 131 -GHEKPSNIKVFIEHLMNGNQLSHMSPTYTTEMGD-ELLAKLADAVEFV 178  
 DB 867 CTTTMSGNMADVLYNLPEIGQVYTNLQFQAYSLGLGDFEFEDVKRYREANILL 918

RESULT 14  
 F86182  
 hypothetical protein [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: F86182  
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719  
 A:Accession: F86182  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1064 <STO>  
 A:Cross-references: GB:AE005172; NID:g7211974; PDB:AAF0445.1; GSPDB:GN00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology

Query Match 6.6%; Score 87; DB 2; Length 1064;  
 Best Local Similarity 21.1%; Pred. No. 23;  
 Matches 60; Conservative 33; Mismatches 103; Indels 88; Gaps 12;  
 QY 6 HWIPVAVSL-----GLLAGSSASAAEEAF-----DLWNEC----- 36  
 DB 601 HGLPVAATKNGSPVDIHALLNGLLVDPHDEAIALNMLKLVSEKNLWHECRINGWKNIH 660  
 QY 37 -----AKACVLDLKDGVRSRMSVDPALDTNGGVLYHSWYLEGNDALK 82  
 DB 661 LFSWPEHCRTYLTIAAC-----RMRHPQWOTD-----ADEVAADDEFSL-----NDSLK 706  
 QY 83 LAIDNALITSGLTIRLEGVEPNK--PVRYSTYTRQARGSMGLNMLVPIGHEKPSNIKV 140  
 DB 707 DVQDMSLRLSMGDKPSLINGSLEPNSADPVKQIMSRM-----TPLEIKSP----- 752

QY 141 FIEHLMNGNQLSHMSPTYTTEMGDELLAKIARDATFEVRAHESNEMOPTLAISHAGSVY 200  
 DB 753 ---ELQKKQSDNINSGKYPALRRERLVLAPOCYDNEGAPDEKAVPMQ-----NII 803  
 QY 201 MA-OTOPRREKRWSEWASGVLCILDPIDGVYNYLAOQRCLDD 243  
 DB 804 KAVNSDQMAKN-----SGFAISTSMPLDELTRPLKSAKIQVSE 842

RESULT 15  
 S57335  
 cleavage and polyadenylation specificity factor 160k chain - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 05-Nov-1999  
 C:Accession: S57335; S57333  
 R:Jenny, A.; Keller, W.  
 Nucleic Acids Res. 23, 2629-2635, 1995  
 A:Title: Cloning of cDNAs encoding the 160 kDa subunit of the bovine cleavage and pol  
 A:Reference number: S57333; MUID:95380277  
 A:Accession: S57335  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1444 <GEN>  
 A:Cross-references: EMBL:X83097; NID:g953171; PDB:CAA58152.1; PID:g929007  
 A:Accession: S57333  
 A:Molecule type: protein  
 A:Residues: 188-197; 204-216; 403-423; 426-437; 511-519; 573-580; 780-789; 1107-1116; 1163-11

Query Match 6.6%; Score 87; DB 2; Length 1444;  
 Best Local Similarity 25.2%; Pred. No. 35;  
 Matches 61; Conservative 31; Mismatches 92; Indels 58; Gaps 13;  
 QY 9 PLVASLGLLAGSSA--SAAEFADLMNECAKAC-----VLDLK-----DGV 48  
 DB 298 PGVAVLNLSTGTITAFPLRTQDGVNITIDCAQAAFIISYDKNVIISLKGSEIYVLTIDGM 357  
 QY 49 RSSR-MSYDPAIADYNGGVLYHSW-LEGNDALKLAIDNALSI-----TSDGLTIR 99  
 DB 358 RSVRAFHFDKAA-----SVLTTSWVTMEPGYLFSGSRISLTKYTEKLOEPPASTAR 412  
 QY 100 LEGVY--PNKRVRYSTYTRQARGSMGLNMLVPIGHEKPSNIKVFIEHLMNGNQLSHMSPI 157  
 DB 413 EADKKEEPSKRRVDATF-----GWSGSKVP--ODEVDEIFVYSEKOSGTOLA-----T 462  
 QY 158 YTIEMGDELL-----AKIARDATFEVRAHESNEMOPTLA-----SHAGSVYMAQTQ 205  
 DB 463 YSFVVDSTLNLNGPCANAMGEPALFSEFQNSPEPDLIYVCSGYGNALSYLQKSIR 522  
 QY 206 PR 207  
 DB 523 PQ 524

Search completed: November 13, 2001, 14:28:25  
 Job time: 131 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:34:25 ; Search time 17.9 seconds

(without alignments)  
482.256 Million cell updates/sec

Title: US-09-412-558-3

Perfect score: 1323  
Sequence: 1 MHLIPHMLPLVSLGLLAG.....YLAQRCLDDPTWEGKIVRV 252

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SWISSPROT\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1307	98.8	638	1 TOXA_PSEAE	P11439 pseudomonas
2	93.5	7.1	1704	1 ABC3_HUMAN	Q09758 homo sapien
3	89.5	6.8	810	1 MCM6_CAEEL	P34647 caenorhabdl
4	89	6.7	380	1 YC9B_SCPO	009885 schizosacch
5	88	6.7	380	1 I132_HUMAN	Q14627 homo sapien
6	87	6.6	1444	1 CP58_BOVIN	Q10569 bos taurus
7	86.5	6.5	369	1 URE2_HELPY	P14917 heliocobacte
8	86.5	6.5	1040	1 AXOI_HUMAN	Q02246 homo sapien
9	85.5	6.5	695	1 VNDC_MABVP	P35263 marburg vir
10	84.5	6.4	560	1 THSI_HAIVO	Q30561 halobacteri
11	84.5	6.4	692	1 VNDC_MABVM	P27588 marburg vir
12	82	6.2	325	1 THIL_BACSU	Q05514 bacillus su
13	82	6.2	1807	1 ITB4_RAT	Q04632 ratius norv
14	81.5	6.2	281	1 BLAC_STAAU	P00807 staphylococ
15	81	6.1	233	1 PUBB_SALTY	P37449 salmonella
16	81	6.1	401	1 HAT2_YEAST	P39984 saccharomyc
17	81	6.1	412	1 GAST_MOUSE	Q60780 homo sapien
18	81	6.1	421	1 GAST_MOUSE	Q60780 mus musculu
19	81	6.1	532	1 NIFD_METBA	P55170 methanosarc
20	81	6.1	788	1 TRPG_PACH	P25170 p anthranil
21	81	6.1	986	1 CYGR_ARBPV	P15288 arabidopsi
22	79.5	6.0	272	1 THIM_BACSU	P39593 bacillus su
23	79	6.0	361	1 DCAM_HELAN	Q65354 heliantus
24	78.5	5.9	422	1 GAST_RAT	Q55188 ratius norv
25	78.5	5.9	950	1 URBI_USTMA	P40349 ustilago ma
26	78	5.9	226	1 PINT_MOUSE	P23506 mus musculu
27	78	5.9	684	1 RPSD_RHIME	Q59753 rhizobium m
28	78	5.9	780	1 K6PL_HUMAN	P17858 homo sapien
29	78	5.9	848	1 ENV_HV1JR	P20871 human immun
30	77.5	5.9	466	1 AMID_PSEPU	Q69768 pseudomonas
31	77.5	5.9	548	1 VN5I_HAHSV4	Q03068 african hor
32	77.5	5.9	1377	1 RPOC_BORBU	Q51349 borrelia bu
33	77.5	5.9	1442	1 CP5A_HUMAN	Q10570 homo sapien

34	76.5	5.8	614	1 SELB_ECOLI	P14081 escherichia
35	76.5	5.8	622	1 PRIR_HUMAN	P16471 homo sapien
36	76	5.7	149	1 ARGD_BACST	Q07907 bacillus st
37	76	5.7	340	1 TF2D_CAEEL	P32085 caenorhabdl
38	76	5.7	461	1 HMCS_ARATH	P54873 arabidopsid
39	76	5.7	641	1 DXS_RHOCA	P26242 rhodobacter
40	76	5.7	839	1 GYRA_MYCPN	P22446 mycoplasma
41	76	5.7	1088	1 RROB_ROTBR	P17468 bovine rota
42	76	5.7	1706	1 CYAA_BORPE	P15318 bordetella
43	75.5	5.7	416	1 RHLE_PANTR	Q28812 pan troglod
44	75.5	5.7	474	1 FLDB_PSEAE	Q9k3c5 pseudomonas
45	75.5	5.7	503	1 TYPH_METJA	Q58081 methanococc

## ALIGNMENTS

RESULT ID	TOXA_PSEAE	STANDARD	PRT	638 AA.
AC	P11439; 091417;			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	EXOTOXIN A PRECURSOR (NAD-DEPENDENT ADP-RIBOSYLTRANSFERASE (EC 2.4.2.-)).			
GN	ETA OR PA1148.			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
OX	Pseudomonas.			
RN	NCBI_TaxID=287;			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 26-53.			
RX	MDLINE=64194063; PubMed=6201861;			
RA	Gray G.L., Smith D.H., Baldridge J.S., Harkins R.N., Vasil M.L.,			
RA	Chen E.Y., Heyneker H.L.;			
RT	"Cloning, nucleotide sequence, and expression in Escherichia coli of the exotoxin A structural gene of Pseudomonas aeruginosa.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 81:2645-2649(1984).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PA01;			
RX	MDLINE=20437337; PubMed=10984043;			
RA	Strover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,			
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,			
RA	Garber R.L., Goltzy L., Iolentino E., Westbrock-Wadman S., Yuan Y.,			
RA	Brody L.V., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,			
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA	Reizer J., Sailer M.H., Hancock R.E.W., Lozy S., Olson M.V.;			
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";			
RL	Nature 406:959-964(2000).			
RN	[3]			
RP	ACTIVE SITE.			
RX	MDLINE=87250491; PubMed=2885323;			
RA	Carroll S.F., Collier R.J.;			
RT	"Active site of Pseudomonas aeruginosa exotoxin A. Glutamic acid 553 is photolabeled by NAD and shows functional homology with glutamic acid 148 of diphtheria toxin.";			
RL	J. Biol. Chem. 262:8707-8711(1987).			
RN	[4]			
RP	DOMAINS.			
RX	MDLINE=90375493; PubMed=2118903;			
RA	Chaudhary V.K., Jino Y., Gato M.G., Fitzgerald D., Pastan I.;			
RT	"Mutagenesis of Pseudomonas exotoxin in identification of sequences responsible for the animal toxicity.";			
RL	J. Biol. Chem. 265:16306-16310(1990).			
RN	[5]			
RP	DOMAINS.			
RX	MDLINE=91006124; PubMed=2170123;			
RA	Bourdenet S., Vacheron M.-J., Guinand M., Michel G., Arminjon F.;			
RT	"Biochemical and immunochemical studies of proteolytic fragments of exotoxin A from Pseudomonas aeruginosa.";			

RL Eur. J. Biochem. 192:379-385(1990).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 424-638.  
 RX MEDLINE-96016159; PubMed-7568123;  
 RA Li M., Dyda F., Benhar I., Pastan I., Davies D.R.;  
 RT "The crystal structure of Pseudomonas aeruginosa exotoxin III  
 with nicotinamide and AMP: conformational differences with the intact  
 exotoxin.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:9308-9312(1995).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 424-638.  
 RX MEDLINE-96293446; PubMed-8692916;  
 RA Li M., Dyda F., Benhar I., Pastan I., Davies D.R.;  
 RT "Crystal structure of the catalytic domain of Pseudomonas exotoxin A  
 complexed with a nicotinamide adenine dinucleotide analog:  
 implications for the activation process and for ADP-ribosylation.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:6902-6906(1996).  
 CC -1- FUNCTION: THIS TOXIN IS AN NAD-DEPENDENT ADP-RIBOSYLTRANSFERASE.  
 CC IT CATALYZES THE TRANSFER OF THE ADP RIBOSYL MOIETY OF OXIDIZED  
 NAD ONTO ELONGATION FACTOR 2 (EF-2) THUS ARRESTING PROTEIN  
 SYNTHESIS.  
 CC -1- PFM: THE 8 CYSTEINES PARTICIPATE IN INTRACHAIN DISULFIDE BONDS.  
 CC -1- SIMILARITY: REGIONAL SEQUENCE SIMILARITY AT THE ACTIVE SITE  
 WITH DIPHTEHRIA TOXIN (DT).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; A01397; AAB59097.1; -  
 DR EMBL; AE004544; AAC04537.1; -  
 DR PIR; A30347; A30347.  
 DR PDB; 1AER; 10-JUN-96.  
 DR PDB; 1DMA; 15-SEP-95.  
 KW Toxin; Signal; Transferrase; Glycosyltransferase; NAD; 3D-structure.  
 FT SIGNAL 1 25  
 FT CHAIN 26 638  
 FT DOMAIN 26 277  
 FT  
 FT DOMAIN 278 389  
 FT  
 FT DOMAIN 390 429  
 FT DOMAIN 430 638  
 FT ACT\_SITE 465 465  
 FT ACT\_SITE 578 578  
 FT CONFLICT 4 4  
 FT CONFLICT 22 22  
 FT CONFLICT 204 204  
 FT CONFLICT 389 389  
 FT CONFLICT 432 432  
 FT CONFLICT 540 540  
 FT SEQUENCE 638 AA; 69284 MW; 799AD56a27C700A CRC64;  
 SQ  
 Query Match 98.8%; Score 1307; DB 1; Length 638;  
 Best Local Similarity 98.8%; Pred. No. 5.8e-108;  
 Matches 249; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 181 HESNEMPTLAISHAGSYVMAOTOPRERKSEMSGKVLCLDPLDGVNVLAAQRCN 240  
 DB 181 HESNEMPTLAISHAGSYVMAOQPRERKSEMSGKVLCLDPLDGVNVLAAQRCN 240  
 OY 241 LDDTWEGKIYRV 252  
 DB 241 LDDTWEGKIYRV 252  
 RESULT 2  
 ABC3\_HUMAN STANDARD; PRT; 1704 AA.  
 ID ABC3\_HUMAN  
 AC 099758; 092473;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE ATP-BINDING CASSETTE, SUB-FAMILY A, MEMBER 3 (ATP-BINDING CASSETTE  
 DE TRANSPORTER 3) (ATP-BINDING CASSETTE 3) (ABC-C TRANSPORTER).  
 GN ABCA3 OR ABC3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISUE-Thyroid carcinoma;  
 RX MEDLINE-96326608; PubMed-8706931;  
 RA Klinghauer N., Hofmann F.;  
 RT "Primary structure of a novel ABC transporter with a chromosomal  
 RT localization on the band encoding the multidrug resistance-associated  
 RT protein.";  
 RL FEBS Lett. 391:61-65(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-9719225; PubMed-9027511;  
 RA Connors T.D., van Raay T.J., Petry L.R., Klinger K.W., Landes G.M.,  
 RA Burn T.C.;  
 RT "The cloning of a human ABC gene (ABC3) mapping to chromosome  
 RT 16p13.3.";  
 RL Genomics 39:231-234(1997).  
 CC -1- FUNCTION: MAY BE A TRANSPORTER, ITS NATURAL SUBSTRATE HAS NOT BEEN  
 CC FOUND YET (BY SIMILARITY). MAY ACT AS AN EFFLUX PUMP FOR  
 CC CHEMOTHERAPEUTICS DRUGS.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, FOLLOWED BY BRAIN,  
 CC PANCREAS, SKELETAL MUSCLE AND HEART. WEAKLY EXPRESSED IN PLACENTA,  
 CC KIDNEY AND LIVER. ALSO EXPRESSED IN MEDULLARY THYROID CARCINOMA  
 CC CELLS (MTC) AND IN C-CELL CARCINOMA.  
 CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES.  
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN  
 CC ATP BINDING CASSETTE (ABC) DOMAIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY  
 CC (ABC TRANSPORTERS). MDR SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U78735; AAC50967.1; -  
 DR EMBL; X97187; CAA65825.1; -  
 DR MIM; 601615; -  
 DR Interpro: IPR001617; -  
 DR Pfam: PF00005; ABC\_tran; 2.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER; 1.  
 KW ATP-binding; Transport; Transmembrane.  
 FT TRANSMEM 22 42  
 FT TRANSMEM 249 269  
 FT TRANSMEM 307 327  
 FT TRANSMEM 344 364  
 FT POTENTIAL.  
 FT POTENTIAL.  
 FT POTENTIAL.

	FT	TRANSMEM	373	393	POTENTIAL.
	ET	TRANSMEM	405	425	POTENTIAL.
	FT	TRANSMEM	447	467	POTENTIAL.
	FT	TRANSMEM	925	945	POTENTIAL.
	FT	TRANSMEM	1100	1120	POTENTIAL.
	FT	TRANSMEM	1144	1164	POTENTIAL.
	FT	TRANSMEM	1183	1203	POTENTIAL.
	FT	TRANSMEM	1213	1233	POTENTIAL.
	FT	TRANSMEM	1245	1265	POTENTIAL.
	FT	TRANSMEM	1306	1326	POTENTIAL.
	FT	NP_BIND	566	573	ATP (POTENTIAL).
	FT	NP_BIND	1416	1423	ATP (POTENTIAL).
	FT	CONFLICT	36	36	P -> S (IN REF. 2).
	FT	CONFLICT	196	196	L -> P (IN REF. 2).
	SQ	SEQUENCE	1704 AA;	191387 MW;	AFO09BDAF/A04FF5F CRC64;

  

Query Match Best Local Similarity 7.1%; Score 93.5; DB 1; Length 1704;  
Matches 36; Conservative 22; Mismatches 54; Indels 45; Gaps 4;

```

Dy      114 YTRQAGSGMSLNLMLVPIGHERKPSNIKVFH-----ELNAGNSLSMSPITYTEMGEDE 165
           : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db       749 FLKKGYGGAGHYMTLVKEPHCNPEDISOLVVHHHVPAITLESSGALESLTLRRESHNREG 808

OY      166 LLAKIARDATFEVRANSHESNMOPFLTAISHAGVSIV-*****MA 202
           |||
Db       809 LFAKL-----EKKOKELGISAFCSITMEEVFLRVGLVKLYDVSSMDIGAILDP 855

OY      203 OTORPREKRKNSEMASGVLC-LLPRLDDVNYVMIAQR 238
           | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db       856 ALDYOHERRASDMAYDSNLTGCAMPDSPDISGALIEBER 892

```

RESULT 3  
MCHE\_CAEEL STANDARD: PRF: 810 AA.  
AC P34647;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE DNA REPLICATION LICENSING FACTOR MCHE HOMOLOG.  
GN ZK632.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;  
OC Rhabditiidae; Pelodermineae; Caenorhabditis.  
OX NCBI\_Taxid=6239;  
  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks J.,  
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper A., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Tier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thillery-Mieg J., Thomas K., Vaubin M., Vaughan K.,  
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sprat J.,  
RA Woldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans.";  
RL Nature 368:32-38(1994).  
CC - SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
CC - SIMILARITY: BELONGS TO THE MCP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 229095; CAAB2357.1; -.
DR EMBL; 222181; CAAB2357.1; JOINED.
DR EMBL; 222181; CAAB0191.1; -.
DR EMBL; 229095; CAAB0191.1; JOINED.
DR Wormpep; ZK632.1; CE00415.
DR InterPro; IPR001208; -.
DR Pfam; PF00493; MCM; 1.
DR PROSITE; PS00847; MCM.1; 1.
DR PROSITE; PSS0051; MCM.2; 1.
DR Hypothetical protein; Transcription regulation; DNA-binding;
KM Nucleic protein; DNA replication; Cell cycle; ATP-binding.
KW Domain 346 554 MCM.
FT NP_BIND 397 404 ATP (POTENTIAL).
SQ SEQUENCE 810 AA; 91130 MM; 214865CFCF7EEB1 CRC64;

Query Match 6.8%, Score 89.5; DB 1; Length 810;
Best Local Similarity 23.2%; Pred. No. 3.3;
Matches 55; Conservative 39; Mismatches 90; Indels 53; Gaps 13;

QY 17 LAGSSASAAEAA--FDLMNECA---KACVLDLKGVSSSRMSVDPALADNRGQ-----G 66
   | | : : | | | | : : : : : : : : : : |
DB 308 LVGGASAGAVEETDYLDLMSKSTEDRATLTKMSDDKRIEKNIYDSLEPNYIGNHEVRLG 367
   | | : : | | | | : : : : : : : : : : |
QY 67 VLHYMSVEGGDALKLALIDNALSTISDGLFTRLEG-----GVEPNKPVRYSYT 115
   | | : : | | | | : : : : : : : : : : |
DB 368 VL--LMLLG--VAKKSREBTSLRGD--INVCLVGDPTSAKQVLAKEVEESP--RAIYT 420
   | | : : | | | | : : : : : : : : : : |
QY 116 RQARGSWSLNMLVPIGHEKPSNIKYFIHELNAGNLSHMSPIYTIEMGDELLAKLARDAT 175
   | | : : | | | | : : : : : : : : : : |
DB 421 -SGKASSAAGLTAAYVKKDES---FEFVIEAGALMLADNGVCCIDERDKM-----DLK 469
   | | : : | | | | : : : : : : : : : : |
QY 176 FFVRAHESNEMPTLAISHAGYSVYMAOTQPRRKRSEMSAGKVLCLLDPLDGYN 232
   | | : : | | | | : : : : : : : : : : |
DB 470 DOVAIHEAMEQD-TISITKAGYKATLN-----ARASILAANPANGRYD 512
   | | : : | | | | : : : : : : : : : : |

RESULT 4
YC9B_SCHPO
AC Q09885; STANDARD; PRT; 380 AA.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHEITICAL 43.0 KDA PROTEIN C584.11C IN CHROMOSOME III.
GN SPOC584.11C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_Taxid:4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL032824; CAB37424.1; -.
DR Hypothetical protein.
KW SEQUENCE 380 AA; 42961 MM; A65777B672310E68 CRC64;

```

Query Match 6.7%; Score 89; DB 1; Length 380;  
 Best Local Similarity 26.0%; Pred. No. 1.4;  
 Matches 34; Conservative 26; Mismatches 43; Indels 28; Gaps 7;

71 SWALEGNAKLAIDNALSDITSLITRLEGGVPE---NKPVRYSTRQARSGWS---123  
 251 STIVNIGGIAMK---DKIISATVDNMTIEHETLDPTEHMEPRISYEMGKDAEITYE 307  
 124 ---LNLVPIGH-----EKPSNKKFIEHNLGNL--SHMSPI-YTIEMGDELL- 167  
 308 DILSVADPGRRLQRIDVLAIEIPSWLKGFGVSGTRPFIQYSPVKFLKMGDEVIE 367  
 168 --AKLADATF 176  
 368 DEATLENETTE 378

RESULT 5  
 ID 1132 HUMAN STANDARD; PRT; 380 AA.  
 AC 014627; 000667;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE INTERLEUKIN-13 RECEPTOR ALPHA-2 CHAIN PRECURSOR (INTERLEUKIN-13 BINDING PROTEIN).  
 GN IL13RA2 OR IL13R.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NC NCB1\_TaxID=9606;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Renal cell carcinoma;  
 RC MEDLINE=96279273; PubMed=8663118;  
 RA Caput D., Laurent P., Kaghad M., Lelias J.M., Lefort S., Vita N., Ferrara P.;  
 RT Cloning and characterization of a specific interleukin (IL)-13 binding protein structurally related to the IL-5 receptor alpha chain.";  
 RT J. Biol. Chem. 271:16921-16926(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Donaldson D.D., Whitters M.J., Fitz L., Neben T., Finerly H., Henderson S.L., O'Hara R.M., Jr., Turner K.J., Wood C.R., Collins M.;  
 RL Submitted (Oct-1996) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA MEDLINE=97321053; PubMed=9177784;  
 RX Guo J., Apion F., Wellein M.P., Lebeau B., Jacques Y., Minvielle S.;  
 RT "Chromosome mapping and expression of the human interleukin-13 receptor.";  
 RT Genomics 42:141-145(1997).  
 RL GENOMICS 42:141-145(1997).  
 CC -1- FUNCTION: BINDS AS A MONOMER WITH HIGH AFFINITY TO INTERLEUKIN-13 (IL-13), BUT NOT TO IL-4.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X95302; CA64617.1; -  
 CC EMBL: U70981; AAB17170.1; -  
 CC EMBL: Y08768; CAA70021.1; -  
 CC MIM: 300130; -

DR InterPro: IPR001777; -  
 DR InterPro: IPR002465; -  
 DR Pfam: PF00041; In3; 1.  
 DR ProSite: PS01356; HEMATOPO. REC. S.F2; 1.  
 KW Receptor; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 26  
 FT CHAIN 27 380  
 FT DOMAIN 27 343  
 FT TRANSMEM 344 363  
 FT DOMAIN 364 380  
 FT DISULFID 145 155  
 FT DISULFID 184 197  
 FT CARBOHYD 115 115  
 FT CARBOHYD 215 215  
 FT CARBOHYD 290 290  
 FT CARBOHYD 299 299  
 FT SEQUENCE 380 AA; 44176 MW; 3C6ACB1B5562C887 CRC64;

Query Match 6.7%; Score 88; DB 1; Length 380;  
 Best Local Similarity 20.7%; Pred. No. 1.7;  
 Matches 34; Conservative 26; Mismatches 46; Indels 58; Gaps 7;

QY 91 ITSDGLITRLEGGVPEPKPVRYSTRQARSGWSLNLVPIGHKPSNKKFIEHNLGNQ 150  
 DB 223 IRSSYTFEQLONIKVPPLPYLFTFRSCSEIKLKSIPLG---PIPARCFDEIIFRED 279  
 QY 151 LSHM-----SPIYTIEMGDELLAKLADATFFVRAHSENMQPTLAISHAGVYMAQT 204  
 DB 280 DITLVATVENVETITKTNE-----TRLCFYVRS-----KYNLYCSDDGI----- 321  
 QY 205 QPRRKRSEWASGKVLCLDPLDGVNYLAQRCNIDDTWEK 248  
 DB 322 -----WSEW-----SDKOC-----WEGE 334

RESULT 6  
 ID CP5A\_BOVIN STANDARD; PRT; 1444 AA.  
 AC 010569;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR, 160 KDA SUBUNIT (CPSE 160 KDA SUBUNIT)  
 DE Bos taurus (Bovine)  
 OS Bos taurus (Bovine)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NC NCB1\_TaxID=9913;  
 OX [1]  
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RP TISSUE=Thymus;  
 RC MEDLINE=95380277; PubMed=7651824;  
 RA Jenny A., Keller W.;  
 RT "Cloning of cDNAs encoding the 160 kDa subunit of the bovine cleavage and polyadenylation specificity factor.";  
 RT and polyadenylation specificity factor.";  
 RT Nucleic Acids Res. 23:2629-2635(1995).  
 RN [2]  
 RP CHARACTERIZATION.  
 RA MEDLINE=92097544; PubMed=1756731;  
 RX Keller W., Bienroth S., Lang K.M., Christofori G.;  
 RT "Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AUAUAA.";  
 RL EMBO J. 10:4241-4249(1991).  
 CC -1- FUNCTION: CPF PLAYS A KEY ROLE IN PRE-MRNA 3'-END FORMATION, RECOGNIZING THE AUAUAA SIGNAL SEQUENCE AND INTERACTING WITH POLY(A) ADDITION. THIS SUBUNIT IS INVOLVED IN THE RNA RECOGNITION STEP OF THE POLYADENYLATION REACTION.  
 CC -1- SUBUNIT: CPSE IS A HETEROTETRAMER COMPOSED OF FOUR DISTINCT SUBUNITS 160, 100, 70 AND 30 KDA.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOPLASM.



```

CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X63097; CAA58152.1; -
CC NM Nuclear protein: RNA-binding.
CC FT DOMAIN 894 909 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC SO SEQUENCE 1444 AA: 161214 MW: 2268346F9812E0FA CRC64;

Query Match 6.6%; Score 87; DB 1; Length 1444;
Best Local Similarity 25.2%; Pred. No. 12;
Matches 61; Conservative 31; Mismatches 92; Indels 58; Gaps 13;

QY 9 PLVASGLLAGSSA--SAAEAFDLNNECAKAC-----VLDLK-----DGV 48
DB 298 PYGVALLSTTTTAPPLRQEGVRLDCAQAAFIYDKHVISLKGGEIVLTITDGM 357
QY 49 RSSR-MSVDPADADTNGGVLYSMV-LEGNDALKLAINALSI-----TSDGLTIR 99
DB 358 RSVAFHFQDKAAA-----SVLTSTMYTMEPGYLFGLSRLGSLNLTKEKLEPPASTAR 412
QY 100 LEGGVE--PKPVARYSTROARGSMVSLNMLVPIGHEKPSNIKVFIELHNGNOLSHMSPI 157
DB 413 EAADKEPSPKKRVDATT-----GMSGSKSYP--QDEVDEIEVGSSEAGSTQLA---T 462
QY 158 YTEEMGBELL-----AKLADATFEFVAHESNEMOPLAT-----SHAGSVYMAQTQ 205
DB 463 YSEVYCSILNIGPCANAAAGEPAFLSEEPONSPEDLEIVYCSGYKNGALSVLQKSI 522
QY 206 PR 207
DB 523 PQ 524

RESULT 7
URE2_HELPY STANDARD; PRT; 569 AA.
AC P14917:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE UREASE BETA SUBUNIT (EC 3.5.1.5) (UREA AMIDOHYDROLASE).
GN UREB OR HPUB OR HP0072 OR JHP0067.
OS Helicobacter pylori (Campylobacter pylori), and
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_Taxid=210, 85963;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=85P;
RX MEDLINE=91161505; Pubmed=2001995;
RA Labigne A., Cussac V., Courcoux P.;
RA "Shuttle cloning and nucleotide sequences of Helicobacter pylori
RT genes responsible for urease activity.";
RL J. Bacteriol. 173:1920-1931(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CPM630;
RX MEDLINE=90221820; Pubmed=2326167;
RA Clayton C.L., Pallen M.J., Kleanthous H., Wren B.W., Tabagchal S.;
RT "Nucleotide sequence of two genes from Helicobacter pylori encoding
RL Nucleic Acids Res. 18:362-362(1990).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; Pubmed=9252185;
RA Tromb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Cooney J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watney L., Mallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=J99;
RX MEDLINE=99120557; Pubmed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Ulla-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
RN [5]
RP SEQUENCE OF 1-20.
RX MEDLINE=90264448; Pubmed=2188975;
RA Dunn B.E., Campbell G.P., Perez-Perez G.I., Blaser M.J.;
RT "Purification and characterization of urease from Helicobacter
RT pylori.";
RL J. Biol. Chem. 265:9464-9469(1990).
RN [6]
RP SEQUENCE OF 1-15.
RX MEDLINE=90202165; Pubmed=2318539;
RA Hu L.T., Mobley H.L.T.;
RT "Purification and N-terminal analysis of urease from Helicobacter
RT pylori.";
RL Infect. Immun. 58:992-998(1990).
CC -1- CATALYTIC ACTIVITY: UREA + H(2)O = CO(2) + 2 NH(3).
CC -1- COFACTOR: EACH CHAIN BINDS TWO NICKEL IONS.
CC -1- SUBUNIT: CONSIST OF TWO SUBUNITS (ALPHA AND BETA).
CC -1- SIMILARITY: BELONGS TO THE UREASE FAMILY.
CC -1- CAUTION: IN HELICOBACTER THE BETA SUBUNIT IS WHAT IS KNOWN, IN
CC OTHER BACTERIA AS THE ALPHA SUBUNIT.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M60398; AAA25021.1; -
CC DR EMBL: X17079; CAA34933.1; -
CC DR EMBL: A08818; CAA00811.1; -
CC DR EMBL: A07398; CAA00663.1; -
CC DR EMBL: AE000529; AAD07143.1; -
CC DR EMBL: AE001446; AAD05651.1; -
CC DR PIR: S07885; URKCBP.
CC DR PIR: B38537; B38537.
CC DR PIR: B41502; B41502.
CC DR HSSP: P18314; IFWE.
CC DR TIGR: HP0072; -.
CC DR InterPro: IPR001924; -.
CC DR Pfam: PF00449; urease; 1.
CC DR PROSITE: PS00145; UREASE_2; 1.
CC DR PROSITE: PS01120; UREASE_1; 1.
CC DR Hydrolase; Metal-binding; Nickel.
CC KW METAL 136 136 NICKEL 2 (BY SIMILARITY).
CC FT METAL 138 138 NICKEL 2 (BY SIMILARITY).

```

```

FT METAL 219 219 NICKEL 1 AND 2 (BY SIMILARITY).
FT METAL 248 248 NICKEL 1 (BY SIMILARITY).
FT METAL 274 274 NICKEL 1 (BY SIMILARITY).
FT METAL 362 362 NICKEL 2 (BY SIMILARITY).
FT ACT SITE 322 322 BY SIMILARITY.
FT CONFLICT 10 10 V -> A (IN REF. 2).
FT CONFLICT 19 19 D -> A (IN REF. 2).
FT CONFLICT 104 104 M -> T (IN REF. 2).
FT CONFLICT 181 181 W -> F (IN REF. 2).
FT CONFLICT 193 193 L -> F (IN REF. 2).
FT CONFLICT 218 218 F -> Y (IN REF. 2).
FT CONFLICT 273 273 F -> S (IN REF. 2).
FT CONFLICT 540 540 P -> S (IN REF. 2).
FT CONFLICT 554 554 SKRANKVSIAQLFSTF -> LNOPK (IN REF. 2).
SQ SEQUENCE 569 AA; 61683 MW; 4C8A6BC8295584 CRC64;

Query Match 6.5%; Score 86.5; DB 1; Length 569;
Best Local Similarity 25.3%; Pred. No. 3.9;
Matches 41; Conservative 25; Mismatches 57; Indels 39; Gaps 9;

OY 43 DLKGVRSRSPVPAIADTNGCVLHYSVLEGNDALKLIDNALSTIS----- 93
DB 103 DMQDGVKNN-LSVGPATFALAGEGLI-----VTAGG-----IDTHFTSPQOIFPTARA 150
OY 94 DGLTRLEGVEPNKPVRYSTYTRQARGSWSLNMLVPGHEKPSNIKVFIELNAGNQLSH 153
DB 151 SGVTTMIGGCGTGPADGT--NATTTTPGRRNLMKMLRAAEYSNMLG-FLAKGNASNDAS- 206
OY 154 MSPITTEMGDELLAKLARDATFFRAHESNMOPTLAISHA 195
DB 207 -----LADQIEA-----GAIKFIHEDMGTPS-AINHA 234

RESULT 8
AXOL HUMAN STANDARD: PRT: 1040 AA.
ID AXOL HUMAN
AC 002246;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE AXONIN-1 PRECURSOR (AXONAL GLYCOPROTEIN TAG-1) (TRANSIENT AXONAL GLYCOPROTEIN 1).
DE GN TAXI OR TAGI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=93145965; PubMed=8425542;
RA Hasler T.H., Rader C., Stoeckli E.T., Zuehlly R.A., Sonderegger P.;
RT "cDNA cloning, structural features, and eucaryotic expression of human TAG-1/axonin-1."
RL Eur. J. Biochem. 211:329-339(1993).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=94140354; PubMed=8307567;
RA Tsiotra C.P., Karagozios D., Theodorakis K., Michaelidis M.T., Modi W.S., Furlley J.A., Jessel M.T., Papamatheakis J.;
RT "Isolation of the cDNA and chromosomal localization of the gene (TAG1) encoding the human axonal glycoprotein TAG-1."
RL Genomics 18:562-567(1993).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE INITIAL GROWTH AND GUIDANCE OF AXONS. MAY BE INVOLVED IN CELL ADHESION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE NEURONAL MEMBRANE BY A GPI-ANCHOR AND IS ALSO RELEASED FROM NEURONS.
CC -1- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTRAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68274; CAA4835.1; -.
DR EMBL: X67734; CAA47963.1; -.
DR PIR: S28830; S28830.
DR MIM: 190197; -.
DR InterPro: IPR001777; -.
DR InterPro: IPR003006; -.
DR Pfam: PF00041; fn3; 4.
DR Pfam: PF00047; 19; 6.
DR Immunoglobulin domain; Glycoprotein; Signal; GPI-anchor;
KW Cell adhesion; Repeat.
FT CHAIN 1 29 1012
FT PROPE 1013 1040
FT DOMAIN 54 118
FT DOMAIN 148 216
FT DOMAIN 254 313
FT DOMAIN 341 402
FT DOMAIN 433 495
FT DOMAIN 523 594
FT DOMAIN 606 612
FT DOMAIN 611 706
FT DOMAIN 714 809
FT DOMAIN 816 908
FT DOMAIN 917 1003
FT SITE 794 796
FT SITE 794 796
FT CARBOHYD 76 76
FT CARBOHYD 198 198
FT CARBOHYD 204 204
FT CARBOHYD 461 461
FT CARBOHYD 477 477
FT CARBOHYD 498 498
FT CARBOHYD 525 525
FT CARBOHYD 830 830
FT CARBOHYD 918 918
FT CARBOHYD 940 940
FT LIPID 1012 1012
SQ SEQUENCE 1040 AA; 11393 MW; 254E78DD3C28BEF6 CRC64;

Query Match 6.5%; Score 86.5; DB 1; Length 1040;
Best Local Similarity 23.8%; Pred. No. 8.4;
Matches 43; Conservative 29; Mismatches 64; Indels 45; Gaps 8;

OY 25 AAEAFDINMECAKACVLDLKGVRSSRMSVDP-----A1AD 61
DB 444 AAPRAVYVLMKSGEYL-----VNSRVTVPDGTLLIRNISRDEGKYTCFAENEMGK 496
OY 62 TNGGVLYHSMVLEGNDALK--LAIDNALSTISDGLTRLEGVEPNKPVRYSTROAR 119
DB 497 ANSGYGLSV-----RDATKTLAPSSADINIGDNLTLQCHASHDPTMDLFTWTLDD-- 547
OY 120 GSNLNLVPTIGHEKPSNIKVFIELNAGN-QLSHSPITTEMGDELLAKLARDATFFV 178
DB 548 -DFPDDFGGHRKRVNVEETIGDILLNAQLRHGK-YTC-MAQTVVDSASKREFTLV 604
OY 179 R 179
DB 605 R 605

RESULT 9
VNUC_MABVP STANDARD: PRT: 695 AA.
ID VNUC_MABVP
AC P35263;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)

```

DT 30-MAY-2000 (Rel. 39, last annotation update)  
 DE NUCLEOPROTEIN (NUCLEOCAPSID PROTEIN).  
 GN NP.  
 OS Marburg virus (strain Popp).  
 CC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;  
 OC Filovirus.  
 OX NCBI\_TaxID=33728;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96028047; PubMed=7487490;  
 RA Bukreyev A.A., Volchkov V.E., Bilinov V.M., Dryga S.A., Netesov S.V.;  
 RT "The complete nucleotide sequence of the Popp (1967) strain of Marburg  
 virus: a comparison with the Musoke (1980) strain.";  
 RL Arch. Virol. 140:1589-1600(1995).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X68495; CAA48509.1; -  
 DR EMBL; 229337; CAA82536.1; -  
 DR PIR; S32777; S32777.  
 DR PIR; S44049; S44049.  
 KW Nucleoprotein.  
 SQ SEQUENCE 695 AA; 77746 MW; F5DB7B7FC8DE020F CRC64;  
  
 Query Match 6.5%; Score 85.5; DB 1; Length 695;  
 Best Local Similarity 22.7%; Pred. No. 6.2; Mismatches 98; Indels 89; Gaps 14;  
 Matches 65; Conservative 34;

QY 5 PHMIPVASGLLAGSSASAEAFDLMNECAKACVLDKDGVRSSMSVPAIAD--- 61  
 DB 107 PHSPLTALKLTESTESQGRIGLF-----LSFSLPLPKLVGGRASIEKALRQVTV 160  
 QY 62 -----TNQ-----GVLRYSNVL-----GGNDALKLADNLSI 91  
 DB 161 HOEGIVTYENHMLTGMKMYIFGLIRSSFLKFLVLIHOGVNLVTGHDAYDASISNSVGQ 220  
 QY 92 TS-DGLTI-----RLEGVPEPKPVRYSTYRQARSGM--SLMWLVPIGHEKP--- 135  
 DB 221 TRRSGLIVTVLEFIQKIDSGTTLPLVRTSKVKNKNEVAFKALSNLARHGAYAPAR 280  
 QY 136 ----SNIKVFTH---ELNA---GNQLSHMSPIYTIEMGDEL--LAKLRADATF--FYRAH 181  
 DB 281 VLNLSGINNLEHGLYPOLSAIALGVATAHGSTLAGVNVNGEYOOLREAHDAEIKLQRH 340  
 QY 182 ESNMQ-----PTLAIHAGVSVMVAQTOPPREK 210  
 DB 341 EHOEIQAIEDDERKILEQFHQKTEITHSQTLAVLSQ---KREK 383  
  
 RESULT 10  
 ID THS1\_HALVO STANDARD: PRT; 560 AA.  
 AC 030561;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE THERMOSOME SUBUNIT 1 (HEAT SHOCK PROTEIN CCT1).  
 GN CCT1.  
 OS Halobacterium volcanii (Haloflex volcanii).  
 CC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloflex.  
 OX NCBI\_TaxID=2246;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97474254; PubMed=9335278;  
 RA Kuo Y.-P., Thompson D.K., St Jean A., Charlebois R.L., Daniels C.J.;  
 RT "Characterization of two heat shock genes from Haloflex volcanii: a

RT model system for transcription regulation in the Archaea.";  
 RL J. Bacteriol. 179:6318-6324(1997).  
 CC -1- FUNCTION: MOLECULAR CHAPERONE; BINDS UNFOLDED POLYPEPTIDES IN  
 CC VITRO, AND HAS A WEAK ATPASE ACTIVITY (BY SIMILARITY).  
 CC -1- SUBUNIT: FORMS AN OLIGOMERIC COMPLEX OF EIGHT-MEMBERED RINGS  
 CC (BY SIMILARITY).  
 CC -1- SIMILARITY: BY HEAT SHOCK.  
 CC -1- SIMILARITY: BELONGS TO THE TCP-1 CHAPERONIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF010470; AAB81497.1; -  
 DR HSSP; P48424; JASX.  
 DR InterPro; IPR002423; -  
 DR Pfam; PF00118; cpn60\_TCP1; 1.  
 DR PRINTS; PR00304; TCOMPLEXTCP1.  
 DR PROSITE; PS00750; TCP1\_1; 1.  
 DR PROSITE; PS00751; TCP1\_2; 1.  
 DR PROSITE; PS00995; TCP1\_3; 1.  
 KW Chaperone; ATP-binding; Multigene family; Heat shock.  
 SQ SEQUENCE 560 AA; 58925 MW; 88B73B2AD70DC341 CRC64;  
  
 Query Match 6.4%; Score 84.5; DB 1; Length 560;  
 Best Local Similarity 19.2%; Pred. No. 5.8;  
 Matches 43; Conservative 51; Mismatches 75; Indels 55; Gaps 9;  
 QY 16 LLAGSSASAEAFDLMNECAKACVLDK--DCVRSRMSVPAIADTNGQVLRHSW 73  
 DB 159 IAAATMGKGESEAKDLSELVDVAVLAVDDIDIRNNVSEKVVGTIDNSLVLGVI 218  
 QY 74 L--EGNDALKLADNA--LSITSDGLTIR--LEGV---EPNKPVRYSTYRQARSGMSL 124  
 DB 219 VDKREVDENNPFAVEDANITLIDALEVRETEIDAEVNVVTDPDQLODFLDOEEQLKEMV 278  
 QY 125 NMLVPIGHEKRSNIKVEIHELNAGNQLSHMSPIYTIEMGDELLAKLRADATFYRAHESN 184  
 DB 279 DQLEVEGAD---AVFV---GDGIDDMAOHY-----LAKEGILAVRAKSS 317  
 QY 185 EMQPTLAIHAGVSVMVAQTOPPREKRSSEWASGVLCILDPLD 228  
 DB 318 DL-----KRLARATGGRAVVSIDDIE 338  
  
 RESULT 11  
 ID VNDC\_MABVM STANDARD: PRT; 692 AA.  
 AC P27588;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE NUCLEOPROTEIN (NUCLEOCAPSID PROTEIN).  
 GN NP.  
 OS Marburg virus (strain Musoke).  
 CC Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;  
 OC Filovirus.  
 OX NCBI\_TaxID=33727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92166742; PubMed=1538192;  
 RA Sanchez A., Kiley M.P., Kleink H.D., Feldmann H.;  
 RT "Sequence analysis of the Marburg virus nucleoprotein gene:  
 RT comparison to Ebola virus and other non-segmented negative-strand RNA  
 RT viruses.";  
 RL J. Gen. Virol. 73:347-357(1992).  
 RN [2]  
 RP REVISIONS.

RA Feldmann H.: Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M72714; AAA46563.1; ALT\_SEQ.  
 DR EMBL: Z12132; CAAT8114.1; .  
 DR PIR: J01408; VHIIMV.  
 DR Nucleoprotein.  
 KW SEQUENCE 692 AA; 77456 MW; A0B6FB03CFE0D0F7 CRC64;  
 SQ  
 Query Match 6.4%; Score 84.5; DB 1; Length 692;  
 Best Local Similarity 22.7%; Pred. No. 7.5;  
 Matches 65; Conservative 34; Mismatches 98; Indels 89; Gaps 14;  
 QY 5 PHWIPVASICLAGSSASAEAFDINNECAKACVLDKGVRSRMSVDPADAD-- 61  
 DB 107 PHYSPLIALKLTLESTESORGRIGF-----ISFCSFLPKLVGDRASIEKALQGVY 160  
 QY 62 -----TNGO-----GVLYHSWYLE-----GGNDALKLAIDNALSI 91  
 DB 161 HOEGIVIVYPHMLTGTGMKVIIFGLRSSFTLKFVLIHOGVNLVGTGHDADVSIIISNVQ 220  
 QY 92 TS-DGLTT-----RLGCVPEPNKPVRYSTYRQARGSM--SLMLVPIGHEKP-- 135  
 DB 221 TRFSGLILVKTVEFLIKTGTSGVLIHPVTRSKYKNEVASFKALSNLRHGGYAFAR 280  
 QY 136 -----SNIKVFTH-----ETNA-----GNQLSHMSPIYTIEMGDEL--LAKIARDATF-PYRAH 181  
 DB 281 VNLSTGINNLEHGLXPOLSAIALAVATAGSTLAGVNVGEQYQOLREAHADAEVKLQRRH 340  
 QY 182 ESNEMQ-----PRLAISHAGVSVMNQTPRRRK 210  
 DB 341 EHQEIQAIEDDERKILEQFHLQKTEITHSQTAVLSQ---KREK 383  
 RESULT 12  
 ID THIL\_BACSU STANDARD: PRT: 325 AA.  
 AC 005514;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE THIAMINE-MONOPHOSPHATE KINASE (EC 2.7.4.16) (THIAMINE-PHOSPHATE  
 DE KINASE).  
 GN THIL.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168 / MARBURG;  
 RC MEDLINE=97346038; PubMed=9202461;  
 RA Sadaie Y., Yata K., Fujita M., Sagai H., Itaya M., Kaeshara Y.,  
 RA Ogasawara N.;  
 RT "Nucleotide sequence and analysis of the phbB-rne-groESL region of  
 RT the Bacillus subtilis chromosome.";  
 RL Microbiology 143:1861-1866 (1997).  
 CC -1- CATALYTIC ACTIVITY: ATP + THIAMINE PHOSPHATE = ADP + THIAMINE  
 CC DIPHOSPHATE.  
 CC -1- PATHWAY: THIAMINE BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE THIAMINE-MONOPHOSPHATE KINASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D88802; BAA19714.1; ALT\_INIT.  
 DR EMBL: Z99107; CAB12409.1; .  
 DR Subtilisin; BG10693; thil.  
 DR Thiamine biosynthesis; Transferase; kinase.  
 KW SEQUENCE 325 AA; 35884 MW; 95D2D6762CFDCCB CRC64;  
 SQ  
 Query Match 6.2%; Score 82; DB 1; Length 325;  
 Best Local Similarity 23.1%; Pred. No. 4.8;  
 Matches 56; Conservative 27; Mismatches 81; Indels 78; Gaps 11;  
 QY 39 ACVLDKGVRSRMSVDPADINNGGVLYHSVMEGGNDALKLAIDNALSIISDGLTI 98  
 DB 28 AALYTAHGVQEI-VCDTWEDVHEK--LHYSPEIDIGKALAVNISD-----I 74  
 QY 99 RLEGVEPNKPVRYSTYRQARGSM--SLMLVPIGHEKPSNITKVFIEHINAGNQLSHM-- 154  
 DB 75 AAMGSI-----PKFLVSLANDPSKMT-----ESEIKAMVEGNEELAKLYHMDLIG 119  
 QY 155 -----SPIYTMGDELLAKIAR--DAEFVRAHESNMOTPLAISHAGVS 198  
 DB 120 GDTVSTADKLVTWTVVIGEIERGQACLRSLAKPNDIVF-----VTEIISSAAGLS 170  
 QY 199 VYMAOTOPR-----REKRWSEWASGKVLK-----LLDPIDGV--YNTLAQ 237  
 DB 171 LLEELFNPONSVEVDYTIHKKHREPRVSVGRICSIFFRAALNDVSDGLASELNEIAEA 230  
 QY 238 RC 239  
 DB 231 SC 232  
 RESULT 13  
 ID ITB4\_RAT STANDARD: PRT: 1807 AA.  
 AC 064632;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 38, Last annotation update)  
 DE INTEGRIN BETA-4 PRECURSOR (GP150) (CD104).  
 DE ITGB4.  
 GN Rattus norvegicus (Rat).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Sciatic nerve;  
 RC MEDLINE=97228432; PubMed=9074510;  
 RA Feltz M.L., Arona M., Scherer S.S., Wrabetz L.;  
 RT "Cloning and sequence of the cDNA encoding the beta 4 integrin  
 RT subunit in rat peripheral nerve.";  
 RL Gene 186:299-304 (1997).  
 CC -1- FUNCTION: INTEGRIN ALPHA-6/BETA-4 MAY MEDIATE ADHESIVE AND/OR  
 CC MIGRATORY FUNCTIONS OF EPITHELIAL CELLS (BY SIMILARITY).  
 CC -1- SUBUNIT: DIMER OF AN ALPHA AND BETA SUBUNIT. BETA-4 ASSOCIATES  
 CC WITH ALPHA-6 (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- PPM: THE CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN DISULFIDE  
 CC BONDS.  
 CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way



FT	STRAND	152	152
FT	TURN	165	166
FT	TURN	169	170
FT	STRAND	171	172
FT	HELIX	174	184
FT	TURN	185	187
FT	HELIX	192	203
FT	TURN	204	204
FT	HELIX	206	208
FT	TURN	209	211
FT	HELIX	212	215
FT	TURN	218	219
FT	STRAND	221	228
FT	TURN	232	233
FT	STRAND	235	242
FT	TURN	244	245
FT	STRAND	250	257
FT	TURN	261	262
FT	TURN	267	267
FT	HELIX	268	278
FT	TURN	279	280
SEQ	SEQUENCE	281 AA;	31349 MW; F82A836773C275FE CRC64;



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:34:01 ; Search time 46.47 Seconds  
(without alignments)  
717.471 Million cell updates/sec

Title: US-09-412-558-3  
Perfect score: 1333  
Sequence: 1 MHLIPHWIPVLVASLGLAG.....YLAQQRNLDPTWEKIRV 252

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP. archaea: \*  
2: SP. bacteria: \*  
3: SP. fungi: \*  
4: SP. human: \*  
5: SP. invertebrate: \*  
6: SP. mammal: \*  
7: SP. mhc: \*  
8: SP. organelle: \*  
9: SP. phage: \*  
10: SP. plant: \*  
11: SP. rodent: \*  
12: SP. unclassified: \*  
13: SP. vertebrate: \*  
14: SP. virus: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	7.2	416	2 Q9RJS6	Q9RJS6 streptomyc
2	93.5	7.1	4180	2 Q9ISN6	Q9ISN6 pseudomonas
3	92.5	7.0	220	2 Q9SSW2	Q9SSW2 helicobacte
4	92	7.0	1180	13 Q9DDT1	Q9DDT1 brachydanio
5	90.5	6.8	311	9 Q38139	Q38139 bacteriopho
6	90	6.8	531	11 Q62319	Q62319 mus musculu
7	89	6.7	6873	6 Q28733	Q28733 oryctolagus
8	89	6.7	26926	4 Q10466	Q10466 homo sapien
9	88.5	6.7	220	2 Q9SSW4	Q9SSW4 helicobacte
10	88	6.7	915	5 Q93538	Q93538 caenorhabdi
11	88	6.7	1195	5 Q16921	Q16921 aedes aegypt
12	87.5	6.6	5721	9 Q48490	Q48490 bacteriopho
13	87	6.6	1064	10 Q9MAU0	Q9MAU0 arabidopsis
14	87	6.6	220	2 Q9WXL2	Q9WXL2 helicobacte
15	86.5	6.5	220	2 Q9SSW3	Q9SSW3 helicobacte
16	86.5	6.5	220	2 Q9SSW1	Q9SSW1 helicobacte
17	86.5	6.5	569	2 Q9S005	Q9S005 helicobacte
18	86.5	6.5	573	2 Q9FCD3	Q9FCD3 streptomyc
19	86.5	6.5	573	2 Q9FCD3	Q9FCD3 streptomyc

20	86	6.5	701	5 Q24713	Q24713 drosophila
21	85.5	6.5	408	2 Q9LA06	Q9LA06 lactococcus
22	85.5	6.5	859	14 Q9Q072	Q9Q072 chimpanzee
23	85.5	6.5	2055	11 Q9ZLK3	Q9ZLK3 mus musculus
24	85.5	6.5	2276	5 Q9TX01	Q9TX01 caenorhabdi
25	85	6.4	441	2 P72763	P72763 synecocyst
26	85	6.4	856	14 Q71014	Q71014 human immun
27	85	6.4	856	14 Q70200	Q70200 human immun
28	84.5	6.4	220	2 Q9R717	Q9R717 helicobacte
29	84.5	6.4	559	10 Q23409	Q23409 arabidopsis
30	84.5	6.4	3729	2 Q33956	Q33956 streptomyc
31	84.5	6.4	4587	11 Q9QXA3	Q9QXA3 mus musculu
32	84	6.3	681	10 Q9LV16	Q9LV16 arabidopsis
33	84	6.3	685	2 P95597	P95597 rhizobium e
34	84	6.3	984	2 Q9XR11	Q9XR11 bacillus ce
35	83.5	6.3	344	2 Q9RJR6	Q9RJR6 streptomyc
36	83.5	6.3	406	2 Q9LO15	Q9LO15 streptomyc
37	83.5	6.3	442	2 Q34974	Q34974 bacillus su
38	83	6.3	220	2 Q9ZNC4	Q9ZNC4 helicobacte
39	83	6.3	351	10 Q9LR24	Q9LR24 arabidopsis
40	83	6.3	921	3 Q9HR41	Q9HR41 neurospora
41	83	6.3	1150	2 Q9K9M0	Q9K9M0 bacillus ha
42	83	6.3	9376	2 Q85168	Q85168 pseudomonas
43	82.5	6.2	344	2 Q916T5	Q916T5 pseudomonas
44	82.5	6.2	410	5 Q9NGP6	Q9NGP6 mastigamoeb
45	82.5	6.2	425	2 Q9XCP9	Q9XCP9 ehrlichia s

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	416 AA.
ID	Q9RJS6			
AC	Q9RJS6			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)			
DE	HYPOPHYSEAL 44.1 KDA PROTEIN.			
GN	SCP51.09C.			
OS	Streptomyces coelicolor.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=1902;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2);			
RA	Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;			
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2);			
RA	MEDLINE=97000351; PubMed=8843436;			
RA	Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,			
RA	Kinast H., Hopwood D.A.;			
RT	"A set of ordered cosmids and a detailed genetic and physical map for			
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";			
RL	Mol. Microbiol. 21:77-96(1996).			
DR	EMBL: AL132707; CAB59707.1; -;			
DR	InterPro: IPR002106; -;			
DR	InterPro: IPR002504; -;			
DR	Pfam: PF01513; DUF15; 1.			
DR	PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.			
KW	Hypothetical protein			
SO	SEQUENCE 416 AA; 44059 MW; F57FBFBFB3B54207 CRC64;			

Query Match

7.2%; Score 95; DB 2; Length 416;

Best Local Similarity 24.4%; Pred. No. 1.3; Matches 75; Conservative 33; Mismatches 104; Indels 96; Gaps 17;

```

QY 11 VASIGLAGSSASAEAFDL--W-NECAKACV-LDI--KDVSSSRMSVDP-----I 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 IERYGLVHNGRGCAEAAREVRECMENAVACTIDWDSDGRHAREEDVADGDDPLV 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 ADTNGOGVLYHNVLEGNDALKALD-----NALSITSDS----- 95
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 VTLCGDGTFLRGARLAENDALLGLGVGLTEVPAPARSAIDAVRDGGLPESSM 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 96 -LITR-----LEGVEPKPVRYSTYRQARQ-----SW-----SLMVL 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 LITRARRLEIPREEMALLRY-----GRGPLPPEPRVTDSESGDEMCIALNVTALNDV 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 128 V--PIGHEKPSNIKVFTHLHAGNQLSHMSPITTIEMGDELLAKLARDATFFVRAHESNE 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 VLEKLSRDRQISGVYI---AGRLASYS-----ADALLVATPTGTAISFAAGFPV 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 186 MOP-----TLATSHAGV--SVMAQTOP-----RREKRWSEMAKYLCLDPLDGV 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 VSPRALVFLFAVAPHTMFDSSVYATDEPVGRLILERSGRAVVSIDQLRGVLDDGWL 339
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 YNYTAQQR 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 GYVAPRR 347

RESULT 2
Q915N6 PRELIMINARY; PRT: 4180 AA.
AC Q915N6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE HYPOTHETICAL PROTEIN PA0690.
GN PA0690.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.F., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizler J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
DR EMBL: AE004504; AAG04079.1;
DR InterPro: IPR001969;
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 4180 AA; 430016 MW; EB181EA3E01BC7AC CRC64;

Query Match 7.1%; Score 93.5; DB 2; Length 4180;
Best Local Similarity 27.8%; Pred. No. 58; Matches 50; Conservative 22; Mismatches 81; Indels 27; Gaps 7;

```

```

QY 154 MSPITTIEMG--DELLAKLARDATFFVRAHESNEMOPLAISAGVSVMAQTOPRREKR 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3365 SASLYSGALGGIDPRYDILLRPA-EVRSRDA--FSPPLASTGGLTVAGDTGMFLETR 3421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
Q95SW2 PRELIMINARY; PRT: 220 AA.
ID Q95SW2;
AC Q95SW2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE UREASE (FRAGMENT).
GN UREB.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KP72B;
RA Tanahashi T., Kita M., Kodama T., Sawai N., Yamaoka Y., Mitsuifuji S.,
RA Katoh F., Kashima K., Imanishi J.;
RT "Comparison of PCR-restriction fragment length polymorphism and PCR-
RT direct sequencing method for differentiating Helicobacter pylori ureb
RT gene.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB028036; BAA78628.1;
DR HSSP: P18314; 2KRU.
DR InterPro: IPR001924;
DR Pfam: PD002467;
DR ProDom: PD002467;
DR PROSITE: PS01120; UREASE_1;
FT NON_TER 1
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 220 AA; 22602 MW; 03C30BD6347565C CRC64;

Query Match 7.0%; Score 92.5; DB 2; Length 220;
Best Local Similarity 25.9%; Pred. No. 0.85; Matches 42; Conservative 26; Mismatches 55; Indels 39; Gaps 9;

```

```

QY 43 DLKDVSRSSRMVDPALADTNGQVLYHSMVLEGNDALKLAIDNALSTS----- 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 28 DMQDGYKNN-LSVGPATEALAGEGLI---VTAGG-----IDTHIFISPPQIPIFA 75
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 DGLTILEGVEPKPVRYSTYRQARQSWSLMWLPIGHEKPSNIKVIHLMGNQLSH 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 76 SGVTTMIGGTGPADGT--NATITTPGRNLKMLRAEEYSNMG-FLAKGNASNDAS- 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 154 MSPITTIEMGDELLAKLARDATFFVRAHESNEMOPLAISHA 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 132 -----LADQIEA-----GAIGLKIHESWGTPPS-AINHA 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
Q9DDT1 PRELIMINARY; PRT: 1180 AA.
ID Q9DDT1;
AC Q9DDT1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE PYRUVATE CARBOXYLASE.
GN PC.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Yoder J.A., Litman G.W.;
RT "The zebrafish fh1, slc3a2, men1, pc, fgf3 and cycd1 genes define two

```



```

DE TITIN (FRAGMENT).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxID=9986;
ON [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CE12;
RA Label S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-6805 FROM N.A.
RC STRAIN=CE12;
RA MEDLINE=92258380; PubMed=1582406;
RL Label S., Gautel M., Lakey A., Trinick J.;
RT "Towards a molecular understanding of titin."
RL EMO J. 11:1711-1716(1992).
RN [3]
RP SEQUENCE OF 4305-5320 FROM N.A.
RC TISSUE=PSOAS MUSCLE;
RA MEDLINE=90238553; PubMed=2129545;
RL Label S., Barlow D.P., Gautel M., Gibson T., Holt J., Hsieh C.L.,
RA France U., Leonard K., Marsale J., Whiting A., Trinick J.;
RT "A regular pattern of two types of 100-residue motif in the sequence
of titin."
RL Nature 345:273-276(1990).
CC -1 SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
CC EMBL: X64696; CAA5937.1; -
DR EMBL: X17329; CAA35207.1; -
DR HSSP: P56276; ITLK.
DR InterPro: IPR000282; -
DR InterPro: IPR001777; -
DR InterPro: IPR003006; -
DR Pfam: PF00041; fn3; 50.
DR Pfam: PF00047; ig; 15.
DR PRINTS: PR00014; ENTPEPIL.
DR SMART: SM00060; FN3.1.
KW Muscle protein; Myosin; Repeat.
FT NON_TER 1
FT TER 6875
FT SEQUENCE 6875 AA; 759127 MW; 50C45B84F3668C55 CRC64;
SO
Query Match 6.7%; Score 89; DB 6; Length 6875;
Best Local Similarity 22.1%; Pred. No. 3.3e+02;
Matches 46; Conservative 35; Mismatches 79; Indels 48; Caps 9;
QY 56 DPAIDTNGQVLYHSMVLEGN-DALK-LAIDNALSTISDGLTTLLEGVEPNKVR 111
DB 4917 DAIIIDSTSS---FTSLVDVNRYSKYLTLNSSSGTSAFVTAVLD--TPSPVN 4971
QY 112 YSTRQARSGWSLNLVPIGHEKSPNIKVFHNLNAGNOLSHMSPYITIEGDELLAKLA 171
DB 4972 LKYTEITKDSVSTIWEPL-LDGGSKIKNIYVEKRDSTRKSTAAVVTNCHKSSMKIDQLQ 5030
QY 172 RDAFFFAHESNEM-----QP-----TLAIS-----HAG 196
DB 5031 ECGSYIRVAENYIGLGPARTADPIKVAEVPQPGKITVDVTVNSVLSWTRPEHDG 5090
QY 197 VSVV---MAOTOPRKRKRNSEKASGVYL 221
DB 5091 GSKTIQYIVEMQAKHSEKMECARVNSL 5118

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
ON [1]
RN SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA MEDLINE=96026330; PubMed=7569978;
RL Label S., Kolmer B.;
RT "Titins: giant proteins in charge of muscle ultrastructure and
elasticity."
RL Science 270:293-296(1995).
RN [2]
RP SEQUENCE OF 22277-25376 FROM N.A.
RC MEDLINE=92258380; PubMed=1582406;
RL Label S., Gautel M., Lakey A., Trinick J.;
RT "Towards a molecular understanding of titin."
RL EMO J. 11:1711-1716(1992).
RN [3]
RP SEQUENCE OF 1976-2014 FROM N.A.
RA Label S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RC CHARACTERIZATION.
RX MEDLINE=9531314; PubMed=7607248;
RA Gautel M., Castiglione-Worell M.A., Pfehl M., Motta A., Pastore A.;
RT "A calmodulin-binding sequence in the C-terminus of human cardiac
titin kinase."
RL Eur. J. Biochem. 230:752-759(1995).
CC -1 FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC -1 ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
DISTRIBUTION. DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY
ONE TISSUE. THE SEQUENCE SHOWN HERE IS THAT OF THE HEART ISOFORM
N2-B.
CC -1 TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC -1 SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
KINASES.
CC -1 SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 112
IMMUNOGLOBULIN C2-LIKE DOMAINS AND 132 FIBRONECTIN TYPE III-LIKE
DOMAINS.
CC EMBL: X64698; CAA5939.1; -
DR EMBL: X83270; CAA58243.1; -
DR EMBL: X64697; CAA5938.1; -
DR EMBL: X90568; CAA62188.1; -
DR EMBL: X64699; CAA45940.1; -
DR HSSP: P56276; ITLK.
DR InterPro: IPR000129; -
DR InterPro: IPR000282; -
DR InterPro: IPR000577; -
DR InterPro: IPR000719; -
DR InterPro: IPR001245; -
DR InterPro: IPR001777; -
DR InterPro: IPR002016; -
DR InterPro: IPR003006; -
DR InterPro: IPR003015; -
DR InterPro: IPR003598; -
DR Pfam: PF00041; fn3; 132.
DR Pfam: PF00047; ig; 95.
DR Pfam: PF00069; PKINASE_1.
DR PRINTS: PR00014; ENTPEPIL.
DR PROSITE: PR00726; LEXASERPTASE.
DR PROSITE: PS00933; FGXY_KINASES_1; UNKNOWN_1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1.
DR PROSITE: PSS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PSS0109; PROTEIN_KINASE_TYR; 1.
DR SMART: SM00408; IGC2; 1.
KW Muscle protein; Cytoskeleton; Structural protein; Calmodulin-binding;
Serine/threonine-protein kinase; Alternative splicing; Repeat;

```

Accession	Protein	Domain	Phosphorylation	Site	Site Type	Site Frequency	Site Conservation	Site Conservation Score	Site Conservation Z-score	Site Conservation P-value	Site Conservation Q-value	Site Conservation R-value	Site Conservation S-value	Site Conservation T-value	Site Conservation U-value	Site Conservation V-value	Site Conservation W-value	Site Conservation X-value	Site Conservation Y-value	Site Conservation Z-value	Site Conservation AA	Site Conservation MW	Site Conservation DEDCD	Site Conservation J254D	Site Conservation S5523	Site Conservation CRC64
RW	Immunoglobulin domain	Phosphorylation	4 X 5 AA TANDEM REPEATS OF R-M-S-P-A.	1370	1389																					
FT	DOMAIN		GLU/LYS/PRO/VAL-RICH.	4429	4614																					
FT	DOMAIN		CATALYTIC	24731	25070																					
FT	DOMAIN		CALMODULIN-BINDING.	25030	25056																					
FT	MOD_RES		PHOSPHORYLATION (BY PDPK)	1372	1372																					
FT	MOD_RES		PHOSPHORYLATION (BY PDPK)	1377	1377																					
FT	MOD_RES		PHOSPHORYLATION (BY PDPK)	1382	1382																					
FT	MOD_RES		PHOSPHORYLATION (BY PDPK)	1387	1387																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26171	26171																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26178	26178																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26178	26178																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26190	26190																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26237	26237																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26237	26237																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26249	26249																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26249	26249																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26249	26249																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26249	26249																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26249	26249																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26249	26249																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26249	26249																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26249	26249																					
FT	MOD_RES		PHOSPHORYLATION (POTENTIAL)	26249	26249																					
FT	MOD_RES		PH																							

Query Match	6.7%;	Score 89;	DB 4;	Length 26926;
Best Local Similarity	22.18;	Pred. NO. 2.6e+03;		
Matches 46;	Conservative 34;	Mismatches 80;	Indels 48;	Gaps 9;

[illegible]

RESULT	9			
Q9SSM4				
ID	Q9SSM4	PRELIMINARY;	PRT;	220 AA.
AC	Q9SSM4;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	URASE (FRAGMENT).			
GN	UREB.			
OS	Helicobacter pylori (Campylobacter pylori).			
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;			
CC	Helicobacter.			
NCBI_OX	NCBI_TaxID=210;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KP96B;			
RA	Tanahashi T., Kita M., Kodama T., Sawai N., Yamooka Y., Mitsuftuji S.,			
RA	Katoh F., Kashima K., Imanishi J.,			
RT	"Comparison of PCR-restriction fragment length polymorphism and PCR-			
RT	direct sequencing method for differentiating Helicobacter pylori ureb			
RT	gene.";			
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AB028038; BAA78630.1; -.			
DR	HSSP; P41020; IUBP.			
DR	InterPro: IPR001924; -.			
DR	Pfam: PF00449; urase; 1.			
DR	ProDom: PD002467; -; 1.			
FT	NON_TER	1	1	
FT	NON_TER	220	220	
SEQ	SEQUENCE	220 AA;	22645 MW;	1701B37E2FB21F98 CRC64;

[illegible]

SEQUENCE	10	RESULT
OC93538	PRELIMINARY:	PRT: 915 AA.
ID OC93538		
AC OC93538		
DT 01-FEB-1997 (TREMBlrel. 02, Created)		
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)		
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)		
DE F20D1.6. PROTEIN.		
GN F20D1.6.		
OS Caenorhabditis elegans.		
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;		
OC Rhabditidae; Peloidorinae; Caenorhabditis.		
NCBI_TaxID=6239;		
LN (1)		
RP SEQUENCE FROM N.A.		
RA Burton J.;		
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.		
LN (2)		
RP SEQUENCE FROM N.A.		
RX MEDLINE=94150718; Pubmed=7906399;		
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,		
RA Crofton J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,		
RA Crexton M., Dear S., Du Z., Durbin R., Favellio A., Fulton L.,		
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,		
RA Jones M., Kershaw J., Kirsten J., Laister N., Lattelle P.,		
RA Lighthning J., Lloyd C., Menniray A., Mortimore B., O'Callaghan M.,		
RA Parsons J., Percy C., Riken L., Koopra A., Saunders D., Shownkeen R.,		
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,		
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,		
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;		
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
RT elegans ";		
RL Nature 368:32-38(1994).		
DR EMBL; Z78542; CAB01748.1; -;		
SO SEQUENCE 915 AA; 104584 MW; 0AEAF6A35A58F8566 CRC64;		

Query Match	6.7%	Score 88	DB 5	Length 915
Best Local Similarity	22.7%	Pred No. 19		
Matches 69	Conservative 47	Mismatches 96	Indels 92	Gaps 17

  

QY	6	HWIP	-----VASIGLLAGGSSASA-----	AEEAFDL	32
		:   :	:   :   :   :   :	:   :	
DB	240	HVPLPKOSKYPFNHNYETGLTATGPTSLPDILEIATWRFRENSLTENHSHSDFDI			299
QY	33	-----WN-----EAKAKCVLDLKDGYRS-SMSVDPAI-AADTNGGVLT--HYSWLEEGCNA			80
		:   :	:   :   :   :   :	:   :	
DB	300	NYSSTWNIMKMKSTKSLDVLVDLEILGYVREDPRKALRSDLNCLILGKHYTO-KELPNAF			358
QY	81	LKLAIDNALSI-----TSDGLTIRLEGVEPNKEPVRSYSTRQANGSMLNVLPGIE			133
		:   :   :   :	:   :	:   :	
DB	359	OKLTINSASEVKGIKHAETEDSGT-----NGMPALAMK-----SWNFIENFE			401
QY	134	KRSNIKVFTHLNL-----AGNQLSHMSPIYTIENGDELLAKLARDATFFYRAHSNEM			186
		:   :   :   :	:   :	:   :	
DB	402	ASDSNLSLHLEMLNMAASMEQSDDELVNNVKLTDLTPNKNPSATATFLAPYCAAK			461

```

OY 187 QPT---LAISHGVSVMAGOPRRERK-WSEMAAGKYLCT-----LDPLDGYNYNLAQ 236
DB 462 STYVWMLALANARVMSD-QPRAEPOLWYEF-----LLKREYKEMETVEKYNGIDH 516
OY 237 QRCN 240
DB 517 LQCS 520

RESULT 11
ID Q16921 PRELIMINARY: PRT: 1195 AA.
AC Q16921;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PYRUVATE CARBOXYLASE.
OS Aedes aegypti (yellow fever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Aedes.
OX NCBI_TaxID=7159;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97218699; PubMed=9066123;
RA Tu Z., Hagedorn H.H.;
RT "Biochemical, molecular, and phylogenetic analysis of pyruvate
  carboxylase in the yellow fever mosquito, Aedes aegypti."
RL Insect Biochem. Mol. Biol. 27:133-147(1997).
CC -1- COPACOR: BIOTIN (BY SIMILARITY).
DR EMBL: L36530; AAB64306.1;
DR HSSP: P10802; ITYU.
DR InterPro: IPR000089;
DR InterPro: IPR000891;
DR InterPro: IPR000901;
DR InterPro: IPR001882;
DR InterPro: IPR003379;
DR Pfam: PF00289; CPase_L_chain; 1.
DR Pfam: PF00364; biotin_lipoyl; 1.
DR Pfam: PF00682; HMG_L-like; 1.
DR Pfam: PF02436; PTC_CODA; 1.
DR PROSITE: PS00188; BIOTIN; 1.
DR PROSITE: PS00867; CPASE_2; UNKNOWN_1.
KW Biotin; Pyruvate.
SQ SEQUENCE 1195 AA: 132201 MW: 145583138049BFC1 CRC64:

Query Match 6.7%; Score 88; DB 5; Length 1195;
Best Local Similarity 21.6%; Pred. No. 29;
Matches 50; Conservative 34; Mismatches 84; Indels 64; Gaps 11;

```

```

DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE COMPLETE NUCLEOTIDE SEQUENCE.
OS Bacteriophage Sph1.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10724;
RN [1]
RP SEQUENCE FROM N.A.
RA Alonso J.C., Luder G., Stiege A.C., Chai S., Weise F., Trautner T.A.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: X87918; CAA66538.1;
SQ SEQUENCE 311 AA: 35943 MW: 1BD3468DA07DAF13 CRC64:

```

```

Query Match 6.6%; Score 87.5; DB 9; Length 311;
Best Local Similarity 23.2%; Pred. No. 4.2; Indels 51; Gaps 6;
Matches 35; Conservative 21; Mismatches 44;
OY 52 RMSVDPALAD-TNGGGLVHSVLEGGNDALKLADNALSTSDGLTIRLEGVEPNKPV 110
DB 112 RTNLDGVLDHENGSLT-----IKTASEYLKEBEGEDIPNO-- 150
OY 111 RYSYTRQARGSWSLNMLVPIGHEKPSNIV-----FHELNANGNLSHMSPIYTIEMCD 164
DB 151 -----YMIQVOH-----NIKVEADFAYVVALIGGNKYK-----YIERDD 187
OY 165 ELAKLARDATFFVRAHESNMOPALISHA 195
DB 188 ELISHVGEHYEWNHNSANIPPTASDSA 218

```

```

RESULT 13
ID Q29117 PRELIMINARY: PRT: 572 AA.
AC Q29117; Q29033; Q29034;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE TITIN (CONNECTIN) (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93365278; PubMed=8359022;
RA Fritz J.D., Wolff J.A., Greaser M.L.;
RT "Characterization of a partial cDNA clone encoding porcine skeletal
  muscle titin: comparison with rabbit and mouse skeletal muscle titin
  sequences."
RL Comp. Biochem. Physiol. 105B:357-360(1993).
RN [2]
RP SEQUENCE OF 62-121 AND 181-231 FROM N.A.
RC TISSUE=MUSCLE;
RA Tanabe R., Muroya S., Nakajima I., Chikuni K., Nakai H.;
RT "Skeletal muscle connectin primary structures as related to animal
  species and muscle type."
RL J. Food Sci. 62:451-461(1997).
CC -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
  DOMAIN.
DR EMBL: M97767; AAA02948.1;
DR EMBL: D85840; BAA12876.1;
DR EMBL: D85841; BAA12877.1;
DR InterPro: IPR001777;
DR InterPro: IPR003006;
DR InterPro: IPR003598;
DR Pfam: PF00041; fn3; 3.
DR Pfam: PF00047; 1g; 2.
DR PRINTS: PR00014; FNTYPEIT.
DR SMART: SMO0408; IGC2; 1.
KW Repeat.

```

FT NON\_TER 1 1  
FT NON\_TER 572 572  
SO SEQUENCE 572 AA; 62977 MW; 7423328093F59C9D CRC64;

Query Match  
Best Local Similarity 19.8%; Pred. No. 12;

Matches 41; Conservative 33; Mismatches 81; Indels 52; Gaps 7;

QY 66 GVLHYSNWLEGNDAKLAITNALSTSDGLTIRLEGG-----VE-----PKKPRY 112  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 196 GEIDDAALIDTSSFTSLVDNVRYSKGKYLTLLENSSGTSKSAFVYTESLDTSPVNL 255  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 113 SYTRQARGSMGLNMLVPIGHEKPSNIKVFIEHLNAGNOLSHMSPIYTIEMGDELLAKLAR 172  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 256 KVTETIKDSVSIETWEPPL-LDGGSKIKNIYIEKREATPKSYAAVYTNCHKMSKIDLOE 314  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 173 DATEFVRAHESNEM-----QP-----TLAIS-----HAGV 197  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 315 GCSYFRTVALENEYGIGLPAHTDDPVKVAEVPQPKITVDYTRNSVLSMTKPEHDG 374  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 198 SVV---MAQTOPRERKRWSEWASGKVL 221  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 375 SKIYIYIEMQAKHSEKSECARVKS 401  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

## RESULT 14

Q9MAUO PRELIMINARY; PRT: 1064 AA.  
ID Q9MAUO;  
AC Q9MAUO;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE F13M7.9 PROTEIN.  
GN F13M7.9.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
OC Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Vysotskaya V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Liu, S.,  
RA Li J., Kremetskaia I., Luros J., Araujo R., Au M., Bredel V.,  
RA Buehler E., Conway A., Dewar K., Feng J., Kim C., Kutz D., Li Y.,  
RA Palm C., Shin P., Sun H., Davis R., Ecker J., Federspiel N.,  
RA Theologis A.;  
RT "The sequence of BAC F13M7 from Arabidopsis thaliana chromosome 1.";  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A.;  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Theologis A.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004809; AAF0445.1; -.  
DR InterPro; IPR001296; -.  
DR Pfam; PF00534; Glycos\_transf.1; 1.  
SO SEQUENCE 1064 AA; 119675 MW; 12D97744B1B2BCFE CRC64;

Query Match 6.6%; Score 87; DB 10; Length 1064;  
Best Local Similarity 21.1%; Pred. No. 30;

Matches 60; Conservative 33; Mismatches 103; Indels 88; Gaps 12;

QY 6 HWIPLVASL-----GLLAGSSASAAEAF-----DLWNEC----- 36  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 601 HGLPVAATKNGGPPVDIHRALNGLVDPHDOEALANALKLVSEKNLMHECRINGMKNIH 660  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 37 -----AKACVLDLKDGVRSRMSVDPALDTNGCGVLHSMVLEGNDALK 82  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 661 LFSMPHERCTYLTRIAC-----RMRRHPQWOTD---ADEVAQDDERSL-----NDSLK 706  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 83 LAIDNALSTSDGLTIRLEGGVEPNK--PVRYSTTRQARGSMGLNMLVPIGHEKPSNIKV 140  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 707 DVODMSLRLSMGDKPKSLNGSLFEPSSADPVKQIMSRM-----TPRIKSRP----- 752  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 141 FIEHLNAGNOLSHMSPIYTIEMGDELLAKLARDAFFPRAHESNEMOPTLAISHAGSVY 200  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 753 ---ELQKKQSDNIGSKYPLRRRLRVLLAVDCYDNEGADERKAMVMIQ-----NII 803  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 201 MA-QTOPRERKRWSEWASGKVLCLDPLDGVYNYLAQRCNLD 243  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 804 KAVSSDPQMAKN-----SGFAISTSMPLDELTRFLSKAKIQVSE 842  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

## RESULT 15

Q9WXL2 PRELIMINARY; PRT: 220 AA.  
ID Q9WXL2;  
AC Q9WXL2;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
DE UREASE (FRAGMENT).  
GN UREB.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OX NCBI\_TaxID=210;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=KP48B;  
RA Tanahashi T., Kita M., Kodama T., Sawai N., Yamaoka Y., Mitsuji S.,  
RA Katoh F., Kashima K., Imanishi J.;  
RT "Comparison of PCR-restriction fragment length polymorphism and PCR-  
direct sequencing method for differentiating Helicobacter pylori ureb  
gene";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB028035; BAA78627.1; -.  
DR HSSP; P18314; IFWB.  
DR InterPro; IPR001924; -.  
DR Pfam; PF00449; urease; 1.  
DR PROSITE; PS01120; UREASE\_1; 1.  
FT NON\_TER 1 1  
FT NON\_TER 220 220  
SO SEQUENCE 220 AA; 22672 MW; AD518C4B809686FB CRC64;

Query Match 6.5%; Score 86.5; DB 2; Length 220;  
Best Local Similarity 25.0%; Pred. No. 3.1;  
Matches 45; Conservative 24; Mismatches 66; Indels 45; Gaps 9;

QY 31 DLWNECARAVL-----DLKDVRSRMSVDPALDTNGCGVLHSMVLEGNDALKLA 84  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 10 DIANKDGKIAIGKGNKMDQGVKN--LSVGPATEALAGEGLI-----VTAGG----- 57  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 85 IDNALSTIS-----DGLTIRLEGGVEPNKPVRYSTTRQARGSMGLNMLVPIGHEKP 135  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 58 IDTHIHFTISPOOIPFAFSGVTTMGSGGPADGT--NATTTTPGRRLKMWLRAAEYS 115  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 136 SNIKVFIEHLNAGNOLSHMSPIYTIEMGDELLAKLARDAFFPRAHESNEMOPTLAISHA 195  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
DB 116 MNLGV-----LAGNITSNDASLADQIEAG-----AIGFKIHEDMGTTFS-AINHA 159  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Tue Nov 13 15:33:12 2001

Search completed: November 13, 2001, 14:34:07  
Job time: 433 sec

us-09-412-558-3.rspt



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:27:41 ; Search time 45.32 Seconds

(without alignments)  
16.052 Million cell updates/sec

Title: US-09-412-558-2  
Perfect score: 56  
Sequence: 1 LIGICVAVTVAI 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

A\_Geneseq\_0601:\*  
1: /SIDSR/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
2: /SIDSR/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
3: /SIDSR/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
4: /SIDSR/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
5: /SIDSR/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
6: /SIDSR/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
7: /SIDSR/gcgdata/geneseq/geneseqp/AA1986.DAT:\*  
8: /SIDSR/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
9: /SIDSR/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
10: /SIDSR/gcgdata/geneseq/geneseqp/AA1989.DAT:\*  
11: /SIDSR/gcgdata/geneseq/geneseqp/AA1990.DAT:\*  
12: /SIDSR/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SIDSR/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SIDSR/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SIDSR/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SIDSR/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SIDSR/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SIDSR/gcgdata/geneseq/geneseqp/AA1997.DAT:\*  
19: /SIDSR/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SIDSR/gcgdata/geneseq/geneseqp/AA1999.DAT:\*  
21: /SIDSR/gcgdata/geneseq/geneseqp/AA2000.DAT:\*  
22: /SIDSR/gcgdata/geneseq/geneseqp/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	75.0	553	14	IL-1R. Homo sapie
2	42	75.0	569	10	Derived sequence o
3	42	75.0	569	10	Human Interleukin-
4	42	75.0	569	15	Human Interleukin-
5	42	75.0	569	15	Human IL-1 recepto
6	42	75.0	569	17	Human Interleukin-
7	42	75.0	569	18	Human Interleukin-
8	42	75.0	569	21	Human Interleukin-
9	42	75.0	569	22	A human Interleukin-
10	42	75.0	569	22	Human Interleukin-
11	40	71.4	297	20	Renal cancer assoc

12	40	71.4	298	19	AAW43422
13	39	71.4	380	21	AAB56914
14	39	69.6	325	17	AAR91951
15	39	69.6	325	17	AAR91955
16	39	69.6	354	17	AAR91950
17	39	69.6	354	21	AAV96200
18	37	66.1	20	22	AAB35172
19	37	66.1	139	21	AAG03749
20	37	66.1	320	21	AAG19309
21	37	66.1	320	21	AAG50485
22	37	66.1	404	21	AAG19308
23	37	66.1	404	21	AAG50484
24	37	66.1	408	21	AAG50483
25	37	66.1	409	21	AAG19307
26	37	66.1	470	21	AAG52829
27	36	64.3	74	20	AAV07824
28	36	64.3	111	19	AAW98841
29	36	64.3	118	20	AAV07823
30	36	64.3	125	20	AAV12365
31	36	64.3	166	20	AAV41121
32	36	64.3	273	22	AAB88323
33	36	64.3	344	19	AAB52812
34	36	64.3	422	14	AAR39819
35	36	64.3	453	21	AAV95015
36	36	64.3	453	22	AAB65170
37	36	64.3	455	21	AAV66647
38	36	64.3	2001	22	AAV72549
39	36	64.3	2233	19	AAW48711
40	36	64.3	2233	19	AAW48712
41	36	64.3	2233	19	AAW48713
42	36	64.3	2233	21	AAB08626
43	35	62.5	302	21	AAW47648
44	35	62.5	326	21	AAW47647
45	35	62.5	350	21	AAW47646

#### ALIGNMENTS

RESULT 1	
AAW42060	standard; Protein: 553 AA.
ID	AAW42060;
AC	AAW42060;
XX	
XX	29-APR-1994 (first entry)
XX	
XX	IL-1R.
XX	
XX	Human: tumour necrosis factor receptor; TNF-R; Interleukin-1 receptor;
KM	IL-1R; fusion protein; linker: TNF; IL-1; cachexia; cerebral malaria;
KM	rheumatoid arthritis; diabetes; multiple sclerosis; septic shock;
KM	pulmonary fibrosis; silicosis; allograft; xenograft; rejection;
KM	graft versus host disease; sepsis; inflammation; allergy;
XX	autoimmune dysfunction.
XX	
OS	Homo sapiens.
XX	
XX	Key
XX	Location/Qualifiers
FT	1..20
FT	/note="Signal peptide"
FT	21..553
FT	/note="Mature hll-1R"
FT	337..356
FT	/note="Transmembrane region"
XX	
XX	Region
XX	
XX	WO9319777-A.
XX	
XX	14-OCT-1993.
XX	
XX	26-MAR-1993; 93WO-US02938.
XX	
XX	30-MAR-1992; 92US-0860710.

Rat syntaxin 4 pro  
Human prostate can  
Lung cancer specif  
HCAVIII phosphoty  
Lung cancer specif  
Non-small cell lun  
Human Trif 2 aptam  
Human secreted pro  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Human secreted pro  
H. pylori GPO 154  
Human secreted pro  
Human 5' EST secre  
Vernonia sucrose t  
Human membrane or  
Human induced tumo  
ZmPK1 homologue pr  
Human secreted pro  
Human PRO732 (UNQ3  
Membrane-bound pro  
Human ATP binding  
HPIV-3 US isolate  
HPIV-3 FRN1 cp45 v  
HPIV-3 Vero cp45 v  
Amino acid sequenc  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia

XX (IMMUNEX CORP.  
 PA Smith CA;  
 PI WPI: 1993-336592/42.  
 DR N-PSDB; AAN049933.  
 XX  
 XX New fusion protein tumour necrosis factor and human interleukin-1  
 PT receptor - useful in therapy, diagnosis and assays of e.g.  
 PT rheumatoid arthritis, diabetes, cerebral malaria, sepsis, etc.  
 XX  
 PS Claim 6; Page 61-64; 85pp; English.  
 XX  
 XX The sequences given in AAR42058-59 represent human tumour necrosis  
 CC factor receptor (TNF-R) and the sequences in AAR42060-61 represent  
 CC human interleukin-1 receptor (IL-1R). These sequences were used in  
 CC the production of a fusion protein which conformed to one of the  
 CC formulae:  
 CC TNF-R-linker-TNF-R-linker-IL-1R  
 CC IL-1R-linker-TNF-R-linker-TNF-R or  
 CC TNF-R-linker-TNF-R  
 CC The linker may comprise 5-100 amino acids selected from Gly, Asp,  
 CC Ser, Thr and Ala. These linkers separate the individual moieties  
 CC by such a distance that each component of the fusion protein is  
 CC capable of folding into the secondary or tertiary structure required  
 CC for its biological activity. These fusion proteins may be used in  
 CC therapy, diagnosis and assays for conditions mediated by TNF or IL-1,  
 CC particularly in conditions in which both TNF and IL-1 play a causative  
 CC role. They may be used to treat cachexia, rheumatoid arthritis,  
 CC diabetes, multiple sclerosis, pulmonary fibrosis and silicosis,  
 CC cerebral malaria, allograft and xenograft rejection in graft verses  
 CC host disease, sepsis, septic shock, inflammation, allergies and  
 CC autoimmune dysfunctions.  
 XX  
 SQ Sequence 553 AA;

Query Match 75.0%; Score 42; DB 14; Length 553;  
 Best Local Similarity 66.7%; Pred. No. 14;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LGICVAVTVAI 12  
 :|||||:|  
 Db 322 mgicvltlvtl 333

RESULT 2  
 AAP92001  
 ID AAP92001 standard; protein: 569 AA.  
 XX  
 XX AAP92001:  
 AC  
 XX  
 XX 07-FEB-1990 (first entry)  
 DT  
 XX  
 XX Derived sequence of human interleukin-1 receptor (IL-1R) gene.  
 DE  
 XX  
 XX Human interleukin-1 receptor; IL-1R; human T-cell line clone 22;  
 KM  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Protein 18  
 FT Region 337..356  
 FT  
 XX  
 XX EP381296-A.  
 PN  
 XX  
 XX 31-MAY-1989.  
 PD  
 XX  
 XX 24-NOV-1988; 88EP-0311150.  
 PF  
 XX  
 XX 30-APR-1984; 84US-0605540.  
 PR 21-DEC-1984; 84US-0684560.

XX (IMMUNEX CORP.  
 PA Dower SK, March CJ, Sims JE, Urdal DL;  
 XX  
 PI WPI: 1989-159346/22.  
 DR N-PSDB; AAN90118.  
 XX  
 XX DNA coding for mammalian interleukin-1 receptor  
 PT - used for obtaining protein and antibodies for  
 PT diagnosis and therapy involving immune or inflammatory  
 PT activities  
 XX  
 PS Claim 17; figure 5A-5C; 32pp; English.  
 XX  
 XX Derived sequence of the coding region of a cDNA encoding human IL-1R.  
 CC The protein feature = N-terminus of mature protein.  
 CC The Region feature = transmembrane region.  
 CC  
 XX  
 SQ Sequence 569 AA;

Query Match 75.0%; Score 42; DB 10; Length 569;  
 Best Local Similarity 66.7%; Pred. No. 15;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LGICVAVTVAI 12  
 :|||||:|  
 Db 338 mgicvltlvtl 349

RESULT 3  
 AAP90330  
 ID AAP90330 standard; protein: 569 AA.  
 XX  
 XX AAP90330:  
 AC  
 XX  
 XX 01-NOV-1989 (first entry)  
 DT  
 XX  
 XX Human interleukin-1 receptor.  
 DE  
 XX  
 XX Clone; interleukin-1 receptor; human; immune regulator;  
 KW inflammatory response.  
 XX  
 XX Murine.  
 OS  
 XX  
 XX WO8904838-A.  
 PN  
 XX  
 XX 01-JUN-1989.  
 PD  
 XX  
 XX 04-NOV-1988; 88WO-US03926.  
 PF  
 XX  
 XX 25-NOV-1987; 87US-0258756.  
 PR  
 XX  
 XX (IMMUNEX CORP.  
 PA  
 XX  
 XX Dower SK, March CJ, Sims JE, Urdal DL;  
 PI  
 XX  
 XX WPI: 1989-178365/24.  
 DR N-PSDB; AAN90031.  
 DR  
 XX  
 XX DNA sequences encoding mammalian interleukin-1 receptors  
 PT - used to regulate immune or inflammatory responses or  
 PT detect IL-1 and its receptor interaction  
 PT  
 XX  
 XX Disclosure; Table 4a-c; 51pp; English.  
 PS  
 XX  
 XX Peptide of human interleukin-1 receptor (see AAN90031).  
 CC This is used in human and veterinary medicine to regulate immune  
 CC or inflammatory responses, and to detect and study IL-1  
 CC molecules and their receptors  
 CC  
 XX  
 XX Sequence 569 AA;

Query Match 75.0%; Score 42; DB 10; Length 569;  
 Best Local Similarity 66.7%; Pred. No. 15;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
 :||||:|  
 Db 338 mlgicvltvlvll 349

RESULT 4  
 AAR59090  
 ID AAR59090 standard; Protein: 569 AA.  
 AC AAR59090;  
 DT 27-APR-1995 (first entry)  
 DE Human Interleukin-1 receptor.  
 XX  
 DE human Interleukin-1 receptor; IL-1R; truncated; connective tissue;  
 KM prevention; arthritis; therapy: repairing; regenerating; ligament;  
 KW tendon; cartilage; synovium; prophylaxis; treatment.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 1..335  
 FT /note="extracellular domain"  
 FT Disulfide-bond 44..96  
 FT Disulfide-bond 142..196  
 FT Disulfide-bond 248..312  
 FT Misc-difference 336  
 FT /note="a truncated receptor is generated by  
 mutation of the codon for this residue  
 to a stop codon"  
 FT Region 336..356  
 FT /note="transmembrane region"  
 FT Domain 357..569  
 FT /note="cytoplasmic domain"  
 XX  
 PN WO9420517-A.  
 XX  
 PD 15-SEP-1994.  
 XX  
 PF 07-MAR-1994; 94MO-US02414.  
 XX  
 PR 08-MAR-1993; 93US-0027750.  
 XX  
 PA (UVP1-) UNIV PITTSBURGH.  
 XX  
 PI Bandara G, Evans CH, Gioroso JC, Robbins PD;  
 XX  
 DR WPI: 1994-302952/37.  
 DR N-PSDB; AAQ71901.  
 XX  
 PT Methods for introducing genes into connective tissue cells - for  
 PT treating connective tissue disorders, e.g. arthritis  
 XX  
 XX Example 9; Page 49-51; 88pp; English.  
 CC This protein is the soluble human interleukin-1 receptor (IL-1R),  
 CC by mutating codon 336 (AAG) of the cDNA encoding IL-1R (AAQ71901), to  
 CC form a stop codon (TAG) a truncated receptor, comprising the  
 CC extracellular domain only, is generated. This truncated receptor retains  
 CC its ability to bind interleukin-1 but is released extracellularly and  
 CC therefore is inactive in signal transduction. The methods of the  
 CC invention are useful for preventing the development of arthritis and for  
 CC therapeutic use, eg. for repairing and regenerating the connective  
 CC tissue. The gene is also useful in a compound for parenteral  
 CC administration for prophylaxis or treatment. (Also see AAR59091 for the  
 CC mouse Interleukin-1 receptor).  
 CC

XX  
 SQ Sequence 569 AA;  
 XX  
 Query Match 75.0%; Score 42; DB 15; Length 569;  
 Best Local Similarity 66.7%; Pred. No. 15;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
 :||||:|  
 Db 338 mlgicvltvlvll 349

RESULT 5  
 AAR60870  
 ID AAR60870 standard; Protein: 569 AA.  
 AC AAR60870;  
 DT 11-JUL-1995 (first entry)  
 DE Human IL-1 receptor protein.  
 XX  
 DE Murine; interleukin-1; receptor; glycoprotein; human; IL-1-alpha;  
 KM expression vector; antibody; diagnosis; therapy: IL-1; IL-1R  
 KM regulation; regulate; immune; inflammatory; activity; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 18..569  
 FT /label="mature human IL-1 receptor  
 FT Domain 337..356  
 FT /label="transmembrane domain"  
 XX  
 PN BP623674-A.  
 XX  
 PD 09-NOV-1994.  
 XX  
 PF 24-NOV-1988; 88EP-0311150.  
 XX  
 PR 25-NOV-1987; 87US-0125627.  
 PR 25-FEB-1988; 88US-0160550.  
 XX  
 PA (IMMV ) IMMUNEX CORP.  
 XX  
 PI Dower SK, March CJ, Sims JE, Urdal DL;  
 XX  
 DR WPI: 1994-343308/43.  
 DR N-PSDB; AAQ73764.  
 XX  
 PT Recombinant cell lines expressing interleukin-1 receptor proteins  
 PT - also antibodies and compositions for use in diagnosis and  
 PT therapy  
 XX  
 XX Example 5; Fig 5; 36pp; English.  
 CC The amino acid sequence of the human interleukin-1 (IL-1) receptor  
 CC (IL-1R) encoded by the insert fragment R3A. The gene (AAQ73764) encodes  
 CC a 65 kD protein precursor which is converted to 63.5 kD protein. The  
 CC protein then undergoes glycosylation to produce a protein of around  
 CC 82 kD. The gene was isolated from a cDNA library, derived from RNA from  
 CC a human T-cell line (designated clone 22), by using a probe prepared  
 CC from the murine IL-1R gene sequence (AAQ73762). Nine positive clones  
 CC were isolated and by restriction mapping and sequencing, to reveal the  
 CC complete coding region of the human IL-1R. The purified IL-1 receptor,  
 CC as part of compositions, can be used for the diagnosis of IL-1 or its  
 CC receptors. The receptor will also be useful for the production of  
 CC antibodies used in diagnosis and therapy. In addition, the purified IL-1  
 CC receptor compositions may be used directly to bind or scavenge IL-1,  
 CC thereby providing means for regulating the immune or inflammatory  
 CC activities of IL-1.  
 CC

SQ Sequence 569 AA:

Query Match 75.0%; Score 42; DB 15; Length 569;  
 Best Local Similarity 66.7%; Pred. No. 15;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LGICVAVTVAI 12  
 :|||||:||||  
 Db 338 mlgicvltvlvl 349

## RESULT 6

AA090376  
 ID AAR90376 standard; protein: 569 AA.

AC AAR90376;

DT 22-JUL-1996 (first entry)

DE Human interleukin-1 receptor I-type.

XX Human interleukin-1 receptor I-type; bone resorption inhibitor;  
 KW treatment; bone metabolic disease; osteoporosis; hypercalcaemia.  
 XX

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..20 /note= "signal peptide"

FT Protein 21..569 /note= "mature protein"

FT Protein 21..332 /note= "soluble protein shuIL-1R"

FT Protein 21..339 /note= "soluble protein"

FT Protein

PN EP687470-A2.

PD 20-DEC-1995.

PE 12-JUN-1995; 95EP-0109011.

PR 17-JUN-1994; 94JP-0135275.

XX (FARH ) HOECHST JAPAN LTD.  
 PA (FARH ) HOECHST JAPAN KK.  
 XX

PI Katoch M, Kitamura K;

DR WPI; 1996-041308/05.

XX Composition conty. interleukin-1 receptor, or soluble fragment - for  
 PT treating and prophylaxis of metabolic bone diseases e.g.  
 PT osteoporosis or hypercalcaemia.  
 XX

PS Claim 5; Page 5-7; 10pp; English.

XX Human interleukin-1 receptor I-type (IL-1R) preferably comprises of  
 CC amino acids 1-319, and more preferably is a soluble protein  
 CC comprising of amino acids 1-312. It is useful as a bone resorption  
 CC inhibitor and can be administered intravenously or intramuscularly.  
 CC A pharmaceutical composition, containing IL-1R as an active  
 CC ingredient, helps prevent and treat metabolic diseases of the bone  
 CC which are caused by the abnormal decrease of calcium and bone matrix  
 CC in the bone, e.g. osteoporosis and hypercalcaemia. The IL-1R is a  
 CC superior therapeutic agent for metabolic bone diseases caused by  
 CC more satisfactory effect for metabolic bone diseases caused by  
 CC abnormal bone resorption by inactivating osteoclasts. Bone  
 CC resorption inhibitor activity of the protein has been proved by  
 CC assaying a direct action on osteoclast activity, e.g. by pit  
 CC formation assay.  
 XX

SQ Sequence 569 AA:

Query Match 75.0%; Score 42; DB 17; Length 569;  
 Best Local Similarity 66.7%; Pred. No. 15;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LGICVAVTVAI 12  
 :|||||:||||  
 Db 338 mlgicvltvlvl 349

## RESULT 7

AA06541  
 ID AAM06541 standard; protein: 569 AA.

AC AAM06541;

DT 13-MAR-1997 (first entry)

DE Human interleukin-1 receptor.

XX Interleukin-1 receptor; IL-1R; cartilage; knee joint; chondrocyte;  
 KW synovioocyte; connective tissue; inflammation; arthritis;  
 KW gene therapy; retrovirus.  
 XX

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..17 /label= Sig-Peptide

FT Protein 18..569 /label= Mat-Protein

FT Domain 336..366 /label= Transmembrane.domain

FT Protein

PN W09639196-A1.

PD 12-DEC-1996.

PE 05-JUN-1996; 96WO-US08899.

PR 06-JUN-1995; 95US-0466932.

XX (UYPI-) UNIV PITTSBURGH.  
 PA Evans CH, Glorioso JC, Kang R, Robbins PD;  
 PI WPI; 1997-042874/04.  
 DR N-PSDB; AAT45876.  
 XX

XX Treatment of a mammalian cartilage defect e.g. arthritis - by  
 PT re-implanting chondrocyte(s) and synovioocyte(s) conty. DNA encoding  
 PT a protein to inhibit interleukin-1 associated cartilage damage  
 XX

PS Disclosure; Page 77-79; 112pp; English.

XX The human interleukin-1 receptor (IL-1R) (AAM06541) amino acid  
 CC sequence can be deduced from a cDNA clone (AAT45876) isolated from a  
 CC human T-cell cDNA library. cDNA encoding the extracellular IL-1  
 CC binding domain of IL-1R can be generated by PCR (see also AAT45878-79)  
 CC and incorporated into a retrovirus vector subsequently used to  
 CC transfect a packaging cell line. Viral particles are produced that  
 CC can be used to infect synovial cells in culture. Infected cells  
 CC are then transplanted into the damaged cartilage of a patient.  
 CC Soluble IL-1R is expressed that binds to and neutralises IL-1 in  
 CC vivo, preventing degradation of cartilage, as a means of treating  
 CC arthritis.  
 CC

SQ Sequence 569 AA;

Query Match 75.0%; Score 42; DB 18; Length 569;

Best Local Similarity 66.7%; Pred. No. 15;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LGICVAVTVAI 12  
:|||||:|  
Db 338 mlgicvcltlvl 349

## RESULT 8

AAB37791  
ID AAB37791 standard; Protein; 569 AA.

AC AAB37791;

DT 23-FEB-2001 (first entry)

DE Human Interleukin-1 receptor, type I precursor.

XX Human; Interleukin-1; IL-1; IL-1alpha; IL-1beta; IL-1 receptor;

KW antiinflammatory; haemostatic; antibacterial; immunosuppressive;

KM immunomodulator; cardiant; cytostatic; neuroprotective; respiratory;

KW inflammation; infection; sepsis; cachexia; autoimmune disorder;

KM cardiovascular disorder; chronic myelogenous leukaemia;

KW multiple sclerosis; inflammatory bowel disease; Crohn's disease.

XX Homo sapiens.

OS WO200064479-A1.

PN 02-NOV-2000.

PD 26-APR-2000; 2000WO-US11700.

PF 27-APR-1999; 99US-0301274.

PR (ANTI-) ANTIBODY SYSTEMS INC.

XX Fredeklng TM, Ignatyev GM;

PI WPI: 2000-679646/66.

PT Novel compositions comprising tetracycline or tetracycline-like

XX compounds for the treatment and/or prevention of acute inflammatory

PT responses and diseases, e.g. septic shock and immune complex-induced

XX colitis -

PS Disclosure: Page 150-152; 183pp; English.

XX The present sequence is given in a specification relating to novel

CC compositions and methods containing tetracycline or tetracycline-like

CC compounds for treating and/or preventing acute inflammatory responses and

CC diseases. Such diseases include acute inflammatory conditions associated

CC with viral haemorrhagic diseases (including diseases caused by

CC Bunyviridae, Filoviridae, Flaviviridae or Arenaviridae viruses),

CC parasitic diseases, bacterial infections, sepsis, cachexia, autoimmune

CC disorders, acute cardiovascular events, chronic myelogenous leukaemia and

CC transplanted bone marrow-induced graft-versus-host disease, septic shock,

CC immune complex-induced colitis, cerebrospinal fluid inflammation,

CC multiple sclerosis, inflammatory responses associated with trauma,

CC systemic inflammatory response syndrome (SIRS), adult respiratory

CC distress syndrome (ARDS), acute liver failure, inflammatory bowel disease

CC and Crohn's disease.

XX Sequence 569 AA;

SO

Query Match 75.0%; Score 42; DB 21; Length 569;

Best Local Similarity 66.7%; Pred. No. 15;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LGICVAVTVAI 12  
:|||||:|  
Db 338 mlgicvcltlvl 349

## RESULT 9

AAB30921  
ID AAB30921 standard; Protein; 569 AA.

AC AAB30921;

DT 02-APR-2001 (first entry)

DE A human Interleukin-1 (IL-1) receptor polypeptide.

XX Human; Interleukin-1; leukocyte infiltration; cartilage degradation;

KW IL-1 receptor antagonist; IRAP; tumour necrosis factor-alpha receptor;

KM TNF-alpha receptor; IL-1.

XX Homo sapiens.

OS US6159464-A.

PN 12-DEC-2000.

PD 05-SEP-1997; 97US-0924376.

PF 08-MAR-1993; 93US-0027750.

PR 20-DEC-1990; 90US-0630981.

PR 23-JUL-1996; 96US-0685212.

XX (UYP1-) UNIV PITTSBURGH.

PA Ghalyzani SC, Robbins PD, Evans CH, Giorioso JC;

XX WPI: 2001-090267/10.

DR N-PSDB; AAC86847.

PT Inhibiting leukocyte infiltration or cartilage degradation in mammalian

XX joint, comprises administering a viral vector comprising a nucleotide

PT sequence operably linked to a promoter encoding a protein of interest

XX -

PS Example 9; Columns 47-52; 72pp; English.

XX The present sequence represents a human Interleukin-1 (IL-1) polypeptide.

CC The IL-1 polypeptide is used in the method of the invention. The

CC specification describes a method for inhibiting leukocyte infiltration

CC or cartilage degradation in a joint of a mammal. The method comprises

CC administering a viral vector encoding a protein of interest, such that

CC expression of the protein within the joint results in an inhibition

CC of leukocyte infiltration or cartilage degradation in the joint. This

CC protein is an IL-1 receptor antagonist (IRAP), soluble IL-1 receptor,

CC soluble tumour necrosis factor-alpha (TNF-alpha) receptor, IL-10, or

CC their biologically active fragments or derivatives. The method is used

CC for treating connective tissue disorders by inhibiting leukocyte

XX infiltration and cartilage degradation.

XX Sequence 569 AA;

SO

Query Match 75.0%; Score 42; DB 22; Length 569;

Best Local Similarity 66.7%; Pred. No. 15;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LGICVAVTVAI 12  
:|||||:|  
Db 338 mlgicvcltlvl 349

RESULT 10

AAB59700

ID AAB59700 standard; Protein; 569 AA.

XX

AC AAB59700;

XX

```

DT 27-MAR-2001 (first entry)
XX
XX Human interleukin-1 receptor.
DE Human; interleukin-1 receptor; arthritis; retroviral vector;
XX joint pathology.
KW Homo sapiens.
XX
XX US6156304-A.
PN
XX 05-DEC-2000.
PD
XX
XX 05-SEP-1997; 97US-0924777.
PF
XX 20-DEC-1990; 90US-0630981.
PR 20-OCT-1992; 92US-0963928.
PR 08-MAR-1993; 93US-0027750.
PR 18-JAN-1994; 94US-0183563.
PR 27-JAN-1995; 95US-0381603.
PR 05-DEC-1995; 95US-0567710.
PR 23-JUL-1996; 96US-0685212.
XX
XX (UYPI-) UNIV PITTSBURGH.
PA
XX Glorioso JC, Evans CH, Robbins PD;
PI
XX WPI; 2001-079537/09.
DR N-PSDB; AAF24245.
XX
XX Producing a protein in a mammalian joint, for treating arthritis in a
PT mammalian host, by administering a recombinant vector comprising a
PT nucleic acid encoding the protein
XX
XX Example 3; Fig 8; 60pp; English.
XX
XX The present invention provides a method of producing a protein in the
CC joint of a mammal, involving the intra-articular administration of a
CC viral vector comprising the coding sequence of the desired protein, and
CC where the protein is expressed in a synovial cell. This is useful in the
CC treatment of joint pathologies such as arthritis.
CC
XX Sequence 569 AA:
SQ

```

Query Match 75.0%; Score 42; DB 22; Length 569;  
 Best Local Similarity 66.7%; Pred. No. 15;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 LGICVAVTVAI 12
Db 338 mlgicvltvll 349

```

RESULT 11  
 ID AAY07078 standard; Protein; 297 AA.  
 AC AAY07078;  
 XX  
 DT 02-JUL-1999 (first entry)  
 DE Renal cancer associated antigen precursor sequence.  
 XX  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09904265-A2.  
 XX  
 PD 28-JAN-1999.

```

XX
XX 15-JUL-1998; 98WO-US14679.
PF
XX 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
XX (LUDWIG) LUDWIG INST CANCER RES.
PA
XX Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
XX WPI; 1999-132448/11.
DR
XX New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
XX Disclosure; Page 484; 787pp; English.
XX
XX The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
CC
XX Sequence 297 AA:
SQ

```

Query Match 71.4%; Score 40; DB 20; Length 297;  
 Best Local Similarity 58.3%; Pred. No. 17;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 LGICVAVTVAI 12
Db 275 llaicvsltvll 286

```

RESULT 12  
 ID AAW43422 standard; Protein; 298 AA.  
 AC AAW43422;  
 XX  
 DT 27-APR-1998 (first entry)  
 DE Rat syntaxin 4 protein.  
 XX  
 KW Binding domain; rat; syntaxin; synaptosomal-associated protein; CNS;  
 KW neurotransmitter; presynaptic membrane; central nervous system; tumour;  
 KW neurodegenerative disease; hormonal disorder; immunological disorder.  
 XX  
 OS Rattus sp.  
 XX  
 PN US5693476-A.  
 XX  
 PD 02-DEC-1997.  
 XX  
 PF 24-FEB-1995; 95US-0393985.  
 XX







---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:26:49 : Search time 26.87 seconds  
(without alignments)  
10.050 Million cell updates/sec

Title: US-09-412-558-2  
Perfect score: 56  
Sequence: 1 LIGICAVATVAI 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	75.0	569	1	US-07-821-716-2 Sequence 2, App1
2	42	75.0	569	2	US-08-381-603-2 Sequence 2, App1
3	42	75.0	569	4	US-08-924-376-2 Sequence 2, App1
4	42	75.0	569	5	US-08-685-212-2 Sequence 2, App1
5	42	75.0	569	5	PCT-US94-02414-2 Sequence 2, App1
6	42	75.0	569	5	PCT-US96-08899-2 Sequence 2, App1
7	40	71.4	298	1	US-08-393-985-10 Sequence 10, App1
8	39	69.6	325	1	US-08-276-919-4 Sequence 4, App1
9	39	69.6	325	1	US-08-276-919-13 Sequence 13, App1
10	39	69.6	325	1	US-08-776-088-4 Sequence 4, App1
11	39	69.6	325	5	US-08-776-088-18 Sequence 18, App1
12	39	69.6	325	5	PCT-US95-09145A-4 Sequence 4, App1
13	39	69.6	325	5	PCT-US95-09145A-18 Sequence 18, App1
14	39	69.6	354	1	US-08-276-919-2 Sequence 2, App1
15	39	69.6	354	1	US-08-776-088-2 Sequence 2, App1
16	39	69.6	354	1	US-08-776-088-6 Sequence 6, App1
17	39	69.6	354	1	US-09-325-320-2 Sequence 2, App1
18	39	69.6	354	4	US-09-585-109-2 Sequence 2, App1
19	39	69.6	354	5	PCT-US95-09145A-2 Sequence 2, App1
20	39	69.6	354	5	PCT-US95-09145A-6 Sequence 6, App1
21	36	64.3	344	3	US-09-058-974-1 Sequence 1, App1
22	36	64.3	344	3	US-09-058-974-1 Sequence 1, App1
23	36	64.3	1457	3	US-08-665-259-27 Sequence 27, App1
24	36	64.3	1457	3	US-08-665-259-27 Sequence 27, App1
25	36	64.3	2233	2	US-08-569-853-1 Sequence 1, App1
26	36	64.3	2233	2	US-08-569-853-1 Sequence 1, App1
27	36	64.3	2233	3	US-08-987-439-1 Sequence 1, App1

28	35	62.5	462	3	US-08-788-231A-15 Sequence 15, App1
29	35	62.5	465	3	US-08-788-231A-17 Sequence 17, App1
30	34	60.7	15	1	US-08-467-083-45 Sequence 45, App1
31	34	60.7	15	1	US-08-414-417B-45 Sequence 45, App1
32	34	60.7	15	2	US-08-486-346A-45 Sequence 45, App1
33	34	60.7	15	2	US-08-468-545B-45 Sequence 45, App1
34	34	60.7	15	3	US-08-466-680B-45 Sequence 45, App1
35	34	60.7	51	4	US-08-789-333F-14 Sequence 14, App1
36	34	60.7	51	4	US-09-169-015-24 Sequence 24, App1
37	34	60.7	51	4	US-09-133-94A-14 Sequence 14, App1
38	34	60.7	124	4	US-08-789-333F-41 Sequence 41, App1
39	34	60.7	127	4	US-08-789-333F-43 Sequence 43, App1
40	34	60.7	173	4	US-08-789-333F-42 Sequence 42, App1
41	34	60.7	177	4	US-08-789-333F-44 Sequence 44, App1
42	34	60.7	249	1	US-07-940-605A-10 Sequence 10, App1
43	34	60.7	249	2	US-08-690-096-10 Sequence 10, App1
44	34	60.7	265	2	US-07-857-224B-65 Sequence 65, App1
45	34	60.7	265	2	US-07-857-224B-66 Sequence 66, App1

## ALIGNMENTS

RESULT 1  
US-07-821-716-2  
; Sequence 2, Application US/07821716  
; Patent No. 5319071  
; GENERAL INFORMATION:  
; APPLICANT: Dower, Steven K.  
; APPLICANT: March, Carl J.  
; APPLICANT: Sims, John  
; APPLICANT: Urdal, David L.  
; TITLE OF INVENTION: Soluble Interleukin-1 Receptors  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/821,716  
; FILING DATE: 19920114  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 125627  
; FILING DATE: 25-NOV-1987  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 160550  
; FILING DATE: 25-FEB-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 258756  
; FILING DATE: 13-OCT-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 691551  
; FILING DATE: 25-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wright, Christopher L.  
; REGISTRATION NUMBER: 31,680  
; TELEPHONE/DOCKET NUMBER: 2001-E  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 587-0606  
; INFORMATION FOR SEQ. ID NO.: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 569 amino acids  
; TYPE: AMINO ACID

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-821-716-2

Query Match 75.0%; Score 42; DB 1; Length 569;  
Best Local Similarity 66.7%; Pred. No. 11;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
:|||||:|||||  
DB 338 MIGICVTLTVII 349

RESULT 2  
US-08-381-603-2  
Sequence 2, Application US/08381603  
Patent No. 5838355

GENERAL INFORMATION:

APPLICANT: Glorioso, Joseph C.

APPLICANT: Evans, Christopher H.

TITLE OF INVENTION: Gene Transfer For Treating a Connective Tissue of a

TITLE OF INVENTION: Mammalian Host

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Eckert Seamans Cherin & Mellott

STREET: 1700 Market Street Suite 3232

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/381,603

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Gould, Jr., Lewis F.

REGISTRATION NUMBER: 25,057

REFERENCE/DOCKET NUMBER: 109070-11

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 575-6000

TELEFAX: (215) 575-6015

TELEX: 866172

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 569 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-381-603-2

APPLICANT: Evans, Christopher H.  
APPLICANT: Robbins, Paul D.  
APPLICANT: Bandara, Geethani  
TITLE OF INVENTION: Gene Transfer For Treating a  
TITLE OF INVENTION: Connective Tissue of a Mammalian Host  
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Eckert Seamans Cherin & Mellott

STREET: 1700 Market Street Suite 3232

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/924,376

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/027,750

FILING DATE: 08-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Gould, Jr., Lewis F.

REGISTRATION NUMBER: 25,057

REFERENCE/DOCKET NUMBER: 109070-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 575-6000

TELEFAX: (215) 575-6015

TELEX: 866172

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 569 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-924-376-2

Query Match 75.0%; Score 42; DB 4; Length 569;  
Best Local Similarity 66.7%; Pred. No. 11;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
:|||||:|||||  
DB 338 MIGICVTLTVII 349

RESULT 4  
US-08-685-212-2  
Sequence 2, Application US/08685212  
Patent No. 6228356

GENERAL INFORMATION:

APPLICANT: Glorioso, Joseph C.

APPLICANT: Evans, Christopher H.

TITLE OF INVENTION: Gene Transfer For Treating a  
TITLE OF INVENTION: Connective Tissue of a Mammalian Host  
NUMBER OF SEQUENCES: 6

APPLICANT: Evans, Christopher H.  
APPLICANT: Robbins, Paul D.  
APPLICANT: Bandara, Geethani  
TITLE OF INVENTION: Gene Transfer For Treating a  
TITLE OF INVENTION: Connective Tissue of a Mammalian Host  
NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Eckert Seamans Cherin & Mellott

STREET: 1700 Market Street Suite 3232

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/924,376

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/027,750

FILING DATE: 08-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Gould, Jr., Lewis F.

REGISTRATION NUMBER: 25,057

REFERENCE/DOCKET NUMBER: 109070-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 575-6000

TELEFAX: (215) 575-6015

TELEX: 866172

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 569 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-924-376-2

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/685,212  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/027,750  
FILING DATE: 08-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Jr., Lewis F.  
REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 109070-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6000  
TELEFAX: (215) 575-6015  
TELEX: 866172  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 569 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-685-212-2

Query Match 75.0%; Score 42; DB 4; Length 569;  
Best Local Similarity 66.7%; Pred. No. 11;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
:|||||:| | |  
DB 338 MIGICVTLTVII 349

RESULT 5  
PCT-US94-02414-2  
Sequence 2, Application PC/TUS9402414  
GENERAL INFORMATION:  
APPLICANT: University of Pittsburgh of the Commonwealth System of Higher Education  
TITLE OF INVENTION: Gene Transfer For Treating a Connective Tissue of  
NUMBER OF INVENTION: a Mammalian Host  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Eckert Seamans Cherin & Mellott  
STREET: 1700 Market Street Suite 3232  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/02414  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Jr., Lewis F.  
REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 109070-9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6000  
TELEFAX: (215) 575-6015  
TELEX: 866172  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 569 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

PCT-US94-02414-2  
Query Match 75.0%; Score 42; DB 5; Length 569;  
Best Local Similarity 66.7%; Pred. No. 11;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
:|||||:| | |  
DB 338 MIGICVTLTVII 349

RESULT 6  
PCT-US96-08899-2  
Sequence 2, Application PC/TUS9608899  
GENERAL INFORMATION:  
APPLICANT: University of Pittsburgh of the Commonwealth  
TITLE OF INVENTION: Gene Transfer For Treating a Connective Tissue of a  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Eckert Seamans Cherin & Mellott  
STREET: 1700 Market Street Suite 3232  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/08899  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gould, Jr., Lewis F.  
REGISTRATION NUMBER: 25,057  
REFERENCE/DOCKET NUMBER: 109070-12A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 575-6000  
TELEFAX: (215) 575-6015  
TELEX: 866172  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 569 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-08899-2

Query Match 75.0%; Score 42; DB 5; Length 569;  
Best Local Similarity 66.7%; Pred. No. 11;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
:|||||:| | |  
DB 338 MIGICVTLTVII 349

RESULT 7  
US-08-393-985-10  
Sequence 10, Application US/08393985  
Patent No. 5693476  
GENERAL INFORMATION:  
APPLICANT: Scheller, Richard H.  
TITLE OF INVENTION: Methods and Compositions for Modulation  
of Vesicular Release  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESSES:

ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,985  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholtz, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 8600-0152  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0960  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 298 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-393-985-10

Query Match 71.4%; Score 40; DB 1; Length 298;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
:|||||:|  
Db 275 MIAICVAVTVAI 286

RESULT 8  
US-08-276-919-4  
; Sequence 4, Application US/08276919  
; Patent No. 5589579  
; GENERAL INFORMATION:  
; APPLICANT: Torczynski, Richard M.  
; APPLICANT: Bollon, Arthur P.  
; TITLE OF INVENTION: Lung Cancer Marker  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS  
; STREET: 1201 Elm Street, Suite 4500  
; CITY: Dallas  
; STATE: TX  
; COUNTRY: US  
; ZIP: 75270-2197  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,919  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Neilligan, Mark C.  
; REGISTRATION NUMBER: 36,389  
; REFERENCE/DOCKET NUMBER: B35792  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 214-939-4500  
; TELEFAX: 214-939-4600  
; INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
; LENGTH: 325 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-276-919-4

Query Match 69.6%; Score 39; DB 1; Length 325;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
:|||||:|  
Db 284 ILGICIVVWSI 295

RESULT 9  
US-08-276-919-13  
; Sequence 13, Application US/08276919  
; Patent No. 5589579  
; GENERAL INFORMATION:  
; APPLICANT: Torczynski, Richard M.  
; APPLICANT: Bollon, Arthur P.  
; TITLE OF INVENTION: Lung Cancer Marker  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS  
; STREET: 1201 Elm Street, Suite 4500  
; CITY: Dallas  
; STATE: TX  
; COUNTRY: US  
; ZIP: 75270-2197  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,919  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Neilligan, Mark C.  
; REGISTRATION NUMBER: 36,389  
; REFERENCE/DOCKET NUMBER: B35792  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 214-939-4500  
; TELEFAX: 214-939-4600  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 325 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-276-919-13

Query Match 69.6%; Score 39; DB 1; Length 325;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
:|||||:|  
Db 284 ILGICIVVWSI 295

RESULT 10  
US-08-776-088-4  
; Sequence 4, Application US/08776088  
; Patent No. 5773579  
; GENERAL INFORMATION:  
; APPLICANT: Torczynski, Richard M.

APPLICANT: Bollon, Arthur P.  
TITLE OF INVENTION: Lung Cancer Marker  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SIDLEY & AUSTIN  
STREET: 1201 Elm Street, Suite 4500  
CITY: Dallas  
STATE: TX  
COUNTRY: US  
ZIP: 75270-2197  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,088  
FILING DATE: 19-JUL-95  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Eugenia S. Hansen  
REGISTRATION NUMBER: 31,966  
REFERENCE/DOCKET NUMBER: 10365/05011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 214-981-3300  
TELEFAX: 214-981-3400  
INFORMATION FOR SEQ. ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-776-088-4

Query Match 69.6%; Score 39; DB 1; Length 325;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
DB 284 ILGICIVVSVI 295

RESULT 11  
US-08-776-088-18  
Sequence 18, Application US/08776088  
Patent No. 5773579  
GENERAL INFORMATION:  
APPLICANT: Torczynski, Richard M.  
APPLICANT: Bollon, Arthur P.  
TITLE OF INVENTION: Lung Cancer Marker  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SIDLEY & AUSTIN  
STREET: 1201 Elm Street, Suite 4500  
CITY: Dallas  
STATE: TX  
COUNTRY: US  
ZIP: 75270-2197  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,088  
FILING DATE: 19-JUL-95  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Eugenia S. Hansen  
REGISTRATION NUMBER: 31,966  
REFERENCE/DOCKET NUMBER: 10365/05011

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 214-981-3300  
TELEFAX: 214-981-3400  
INFORMATION FOR SEQ. ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-776-088-18

Query Match 69.6%; Score 39; DB 1; Length 325;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
DB 284 ILGICIVVSVI 295

RESULT 12  
PCT-US95-09145A-4  
Sequence 4, Application PC/TUS9509145A  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Lung Cancer Marker  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS  
STREET: 1201 Elm Street, Suite 4500  
CITY: Dallas  
STATE: TX  
COUNTRY: US  
ZIP: 75270-2197  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/09145A  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: John A. Harre  
REGISTRATION NUMBER: 37,345  
REFERENCE/DOCKET NUMBER: B35792C1PPCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 214-939-4600  
TELEFAX: 214-939-4600  
INFORMATION FOR SEQ. ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 325 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-09145A-4

Query Match 69.6%; Score 39; DB 5; Length 325;  
Best Local Similarity 50.0%; Pred. No. 19;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
DB 284 ILGICIVVSVI 295

RESULT 13  
PCT-US95-09145A-18  
Sequence 18, Application PC/TUS9509145A  
GENERAL INFORMATION:

```

APPLICANT:
TITLE OF INVENTION: Lung Cancer Marker
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: TX
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09145A
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: John A. Harre
REGISTRATION NUMBER: 37,345
REFERENCE/DOCKET NUMBER: B35792CIPPT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 325 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-09145A-18

Query Match          69.6%; Score 39; DB 5; Length 325;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVYVAI 12
Db 284 ILGICIVVYSI 295

RESULT 14
US-08-276-919-2
; Sequence 2, Application US/08276919
; Patent No. 5589579
; GENERAL INFORMATION:
; APPLICANT: Torczynski, Richard M.
; APPLICANT: Bollon, Arthur P.
; TITLE OF INVENTION: Lung Cancer Marker
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: TX
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,919
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Neilligan, Mark C.
; REGISTRATION NUMBER: 36,389
; REFERENCE/DOCKET NUMBER: B35792

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-276-919-2

Query Match          69.6%; Score 39; DB 1; Length 354;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVYVAI 12
Db 313 ILGICIVVYSI 324

RESULT 15
US-08-776-088-2
; Sequence 2, Application US/08776088
; Patent No. 5773579
; GENERAL INFORMATION:
; APPLICANT: Torczynski, Richard M.
; APPLICANT: Bollon, Arthur P.
; TITLE OF INVENTION: Lung Cancer Marker
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SIDLEY & AUSTIN
; STREET: 1201 Elm Street, Suite 4500
; CITY: Dallas
; STATE: TX
; COUNTRY: US
; ZIP: 75270-2197
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/776,088
; FILING DATE: 19-JUL-95
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Eugenia S. Hansen
; REGISTRATION NUMBER: 31,966
; REFERENCE/DOCKET NUMBER: 10365/05011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 214-981-3300
; TELEFAX: 214-981-3400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-776-088-2

Query Match          69.6%; Score 39; DB 1; Length 354;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVYVAI 12
Db 313 ILGICIVVYSI 324

```

Search completed: November 13, 2001, 14:26:50



Job time: 36 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:28:18 ; Search time 30.29 Seconds  
(without alignments)  
30.178 Million cell updates/sec

Title: US-09-412-558-2

Perfect score: 56

Sequence: 1 LIGICVAVTVAI 12

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR\_68:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	100.0	68	2 E36849	A13L protein - var
2	56	100.0	68	2 T28555	hypothetical prote
3	56	100.0	70	2 F42518	A13L protein - vac
4	56	100.0	70	2 T37400	structural protein
5	52	92.9	68	2 C72165	A14L protein - var
6	42	75.0	569	2 A36187	interleukin-1 rece
7	40	71.4	237	2 C64637	amino acid ABC tra
8	40	71.4	287	2 T38517	synthaxin - human
9	40	71.4	287	2 S52726	synthaxin-4 - human
10	40	71.4	298	2 E48213	synthaxin 4 - rat
11	38.5	68.8	573	2 T19880	hypothetical prote
12	38	67.9	114	2 T43080	hypothetical prote
13	38	67.9	207	2 T32331	hypothetical prote
14	38	67.9	267	2 B83705	phosphonates trans
15	38	67.9	281	2 D70729	probable peptide t
16	37	67.9	353	2 T24616	hypothetical prote
17	37	66.1	135	2 G83495	hypothetical prote
18	37	66.1	141	1 HAOMP	hemoglobin alpha c
19	37	66.1	193	2 S70681	bpik protein - Bor
20	37	66.1	218	2 E69843	conserved hypothet
21	37	66.1	359	2 H70579	probable murex prot
22	37	66.1	415	2 A86056	low affinity trypit
23	37	66.1	415	2 A39412	tryptophan transpo
24	37	66.1	436	2 C84462	hypothetical prote
25	37	66.1	599	2 G70713	hypothetical prote
26	37	66.1	973	2 T41272	hypothetical prote
27	36	64.3	111	2 A64563	hypothetical prote
28	36	64.3	141	1 HAIGI	hemoglobin alpha-I
29	36	64.3	350	2 I46608	MHC PD6-glycoprote

30	36	64.3	364	2 G86340	protein F2D10.35 (
31	36	64.3	366	2 T27257	hypothetical prote
32	36	64.3	422	2 S52578	serine/threonine-s
33	36	64.3	457	2 T46332	hypothetical prote
34	36	64.3	492	2 T02458	hypothetical prote
35	36	64.3	1472	2 B54774	ATP-binding casset
36	36	64.3	1529	2 A59189	ATP-binding casset
37	36	64.3	2233	1 ZLN2P3	genome polyprotein
38	35	62.5	134	2 I46090	ubiquinol--cytochr
39	35	62.5	134	2 I46091	ubiquinol--cytochr
40	35	62.5	141	2 A61485	hemoglobin alpha c
41	35	62.5	157	2 S57997	probable olfactory
42	35	62.5	170	2 S56958	probable membrane
43	35	62.5	184	2 B72168	A37R protein - var
44	35	62.5	184	2 A36852	A36R protein - var
45	35	62.5	184	2 T28578	BR protein - vario

#### ALIGNMENTS

RESULT 1  
E36849  
A13L protein - variola virus (strain India-1967)  
C:Species: Variola virus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 23-Mar-2001  
C:Accession: E36849  
R:Binov, V.M.  
submitted to GenBank, November 1992  
A:Reference number: A36859  
A:Accession: E36849  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-68 <BLU>  
A:Cross-references: GB:X69198; NID:g456758; PIDN:CAA49058.1; PID:g297296

Query Match 100.0%; Score 56; DB 2; Length 68;

Best Local Similarity 100.0%; Pred. No. 0.0083; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
|||||  
Db 7 LIGICVAVTVAI 18

#### RESULT 2

T28555  
hypothetical protein A14L - variola major virus  
C:Species: variola major virus  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T28555

R:Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Au Nature 366, 748-751, 1993  
A:Title: Potential virulence determinants in terminal regions of variola smallpox vir

A:Reference number: Z20488; MUID:94088747

A:Accession: T28555

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-68 <MAS>

A:Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60865.1; PID:g439035

A:Experimental source: strain Bangladesh-1975

Query Match 100.0%; Score 56; DB 2; Length 68;

Best Local Similarity 100.0%; Pred. No. 0.0083; Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
|||||  
Db 7 LIGICVAVTVAI 18

RESULT 3  
F42518  
A:13L protein - vaccinia virus (strain Copenhagen)  
C:Species: vaccinia virus  
A:Note: host Homo sapiens (man)  
C:Date: 09-Nov-1990 #sequence\_revision 09-Nov-1990 #text\_change 08-Apr-1994  
C:Accession: F42518  
R:Johnson, G. P.  
Submitted to GenBank, June 1990  
A:Reference number: A33172  
A:Accession: F42518  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-70 <JCH>

Query Match 100.0%; Score 56; DB 2; Length 70;  
Best Local Similarity 100.0%; Pred. No. 0.0085;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
|||||  
Db 7 LIGICVAVTVAI 18

RESULT 4  
T37400  
structural protein p8 - vaccinia virus (strain Ankara)  
C:Species: vaccinia virus  
A:Variety: strain Ankara  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
C:Accession: T37400  
R:Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F.  
Submitted to the EMBL Data Library, March 1997  
A:Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) strain  
A:Reference number: Z20877  
A:Accession: T37400  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-70 <ANT>  
A:Cross-references: EMBL:U94848; PIDN:AAB96464.1  
A:Experimental source: strain Ankara  
C:Genetics:  
A:Note: MVA124L

Query Match 100.0%; Score 56; DB 2; Length 70;  
Best Local Similarity 100.0%; Pred. No. 0.0085;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
|||||  
Db 7 LIGICVAVTVAI 18

RESULT 5  
C72165  
A14L protein - variola minor virus (strain Garcia-1966)  
C:Species: variola minor virus  
C:Date: 24-Nov-1999 #sequence\_revision 24-Nov-1999 #text\_change 24-Nov-1999  
C:Accession: C72165  
R:Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Saitonov, P.F.; Massung, R.F.; Lopar  
Submitted to GenBank, March 1998  
A:Description: Analysis of the complete coding sequence of DNA of alastrim variola minor  
A:Reference number: A72150  
A:Accession: C72165  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-66 <SHC>  
A:Cross-references: GB:Y16780; NID:G5830555; PIDN:CAB54717.1; PID:61542673; PID:G5830678  
A:Experimental source: strain Garcia-1966  
C:Genetics:  
A:Gene: A14L

Query Match 92.9%; Score 52; DB 2; Length 68;  
Best Local Similarity 91.7%; Pred. No. 0.039;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
|||||  
Db 7 LIGICVAVTVAI 18

RESULT 6  
A36187  
Interleukin-1 receptor type I precursor - human  
C:Species: Homo sapiens (man)  
C:Date: 14-Dec-1990 #sequence\_revision 14-Dec-1990 #text\_change 29-Sep-1999  
C:Accession: A36187; S06928  
R:Sim, J.E.; Acres, R.B.; Grubbs, C.E.; McMahon, C.J.; Wignall, J.M.; March, C.J.; D  
Proc. Natl. Acad. Sci. U.S.A. 86, 8946-8950, 1989  
A:Title: Cloning the Interleukin 1 receptor from human T cells.  
A:Reference number: A36187; MUID:90046906  
A:Accession: A36187  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-569 <SIM>  
A:Cross-references: GB:M20658  
R:Chua, A.O.; Gubler, U.  
Nucleic Acids Res. 17, 10114, 1989  
A:Title: Sequence of the cDNA for the human fibroblast type Interleukin-1 receptor.  
A:Reference number: S06928; MUID:90098789  
A:Accession: S06928  
A:Status: translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-569 <CHU>  
A:Cross-references: EMBL:X16896; NID:G33800; PIDN:CAA34773.1; PID:G33801  
C:Genetics:  
A:Gene: GDB:IL1R1; IL1RA; D2S1473; IL1R  
A:Cross-references: GDB:125254; OMIM:147810  
A:Map position: 2q12-2q12  
C:Superfamily: Interleukin-1 receptor type I  
C:Keywords: cytokine receptor; transmembrane protein

Query Match 75.0%; Score 42; DB 2; Length 569;  
Best Local Similarity 66.7%; Pred. No. 11;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
|||||  
Db 338 MIGICVTLTVII 349

RESULT 7  
C64637  
amino acid ABC transporter, permease protein - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori  
C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 20-Aug-1999  
C:Accession: C64637  
R:Tom, J.F.; White, O.; Kerlavase, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khairi, H.G.; Glodek, A.; McK  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Wathey,  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karik, P.D.; Smith, H.O.; Fraser,  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467  
A:Accession: C64637  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-237 <TOM>  
A:Cross-references: GB:AE000603; GB:AE000511; NID:G2314075; PIDN:AAD07984.1; PID:G231  
C:Superfamily: histidine permease protein M

Query Match 71.4%; Score 40; DB 2; Length 237;  
Best Local Similarity 50.0%; Pred. No. 12;  
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 LIGICAVATVAI 12  
:|||||:  
Db 48 ILGICIAVVAL 59

## RESULT 8

I38517

syntaxin - human

C:Species: Homo sapiens (man)

C&gt;Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 24-Sep-1999

C:Accession: I38517

R:Li, H.; Hodge, D.R.; Pel, G.K.; Seth, A.

Gene 143, 303-304, 1994

A&gt;Title: Isolation and sequence analysis of the human syntaxin-encoding gene.

A:Reference number: I38517; MUID:94266173

A:Accession: I38517

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-297 &lt;RES&gt;

A:Cross-references: EMBL:U07158; NID:9463906; PIDN:AAA20967.1; PID:9463907

C:Superfamily: syntaxin

Query Match 71.4%; Score 40; DB 2; Length 297;  
Best Local Similarity 58.3%; Pred. No. 14;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICAVATVAI 12  
|||||:  
Db 275 LIAICVSITVVL 286

## RESULT 9

S52726

syntaxin-4 - human

C:Species: Homo sapiens (man)

C&gt;Date: 19-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 24-Sep-1999

C:Accession: S52726

R:Jagadiish, M.N.; Fernandez, C.S.

submitted to the EMBL Data Library, December 1994

A&gt;Description: Isolation and sequence analysis of the syntaxin-4 encoding region from hu

A:Reference number: S52726

A:Accession: S52726

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-297 &lt;JAG&gt;

A:Cross-references: EMBL:X85784; NID:9758104; PIDN:CAA59769.1; PID:9758105

C:Superfamily: syntaxin

Query Match 71.4%; Score 40; DB 2; Length 297;  
Best Local Similarity 58.3%; Pred. No. 14;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICAVATVAI 12  
|||||:  
Db 275 LIAICVSITVVL 286

## RESULT 10

E48213

syntaxin 4 - rat

C:Species: Rattus norvegicus (Norway rat)

C&gt;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 24-Sep-1999

C:Accession: E48213

R:Bennett, M.K.; Garcia-Arrears, J.E.; Elferink, L.A.; Peterson, K.; Fleming, A.M.; Hazu

Cell 74, 863-873, 1993

A&gt;Title: The syntaxin family of vesicular transport receptors.

A:Reference number: A48213; MUID:93386759

A:Accession: E48213  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-298 <BEN>  
A:Cross-references: GB:I20821; NID:9349320; PIDN:AAA03046.1; PID:9349321  
C:Superfamily: syntaxin

Query Match 71.4%; Score 40; DB 2; Length 298;  
Best Local Similarity 66.7%; Pred. No. 14;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICAVATVAI 12  
:|||||:  
Db 275 MIAICSVITVLI 286

## RESULT 11

T19880

hypothetical protein C41G6.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T19880

R:Cummings, P.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19191

A:Accession: T19880

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-573 &lt;WIL&gt;

A:Cross-references: EMBL:Z81047; PIDN:CAB02826.1; GSPDB:GN00023; CESP:C41G6.2

C:Genetics:

A:Gene: CESP:C41G6.2

A:Map position: 5

A:Introns: 203/2; 305/1; 472/2

Query Match 68.8%; Score 38.5; DB 2; Length 573;  
Best Local Similarity 61.5%; Pred. No. 44;  
Matches 8; Conservative 4; Mismatches 0; Indels 1; Gaps 1;

QY 1 LIGICVA-VTVAI 12  
|||||:  
Db 76 MIGICADLITAI 88

## RESULT 12

T43080

hypothetical protein - Lactococcus lactis plasmid pmRC01

C:Species: Lactococcus lactis

C&gt;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jun-2000

C:Accession: T43080

R:Dougherty, B.A.; Hill, C.; Weidman, J.F.; Richardson, D.R.; Venter, J.C.; Ross, R.P

Mol. Microbiol. 29, 1029-1038, 1998

A&gt;Title: Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasmid

A:Reference number: Z22314

A:Accession: T43080

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-114 &lt;DOU&gt;

A:Cross-references: EMBL:AE001272; PIDN:AAC55994.1

A:Experimental source: strain DPC3147

C:Genetics:

A:Genome: plasmid pmRC01

A:Note: ORF0008

C:Superfamily: Staphylococcus aureus transfer complex protein trsb

Query Match 67.9%; Score 38; DB 2; Length 114;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```
QY      1 LGICVAITVAI 12  
         |:|:|:|:|:|  
Db      59 LVGVCAVAFIII 70
```

## RESULT 13

hypothetical protein C24H12.8 - *Caenorhabditis elegans*  
C.Species: *Caenorhabditis elegans*  
C.Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C.Accession: J32331  
R.Rohlfing, T.; Wohlmann, P.; Biewald, T.  
submitted to the EMBL Data Library, September 1997  
A.Description: The sequence of C. elegans cosmid C24H12.  
A.Reference number: Z1151  
A.Accession: J32331  
A.Status: preliminary; translated from GB/EMBL/DDBB  
A.Molecule type: DNA  
A.Residues: 1-207 <RDB>  
A.Cross-references: EMBL:AF025451; PIDN:AA871204.1; GSPDB:GMD0020; CESP:C24H12.8  
A.Experimental source: strain Bristol N2; clone C24H12  
C.Genetic: CESP:C24H12.8  
A.gene: CESP:C24H12.8  
A:Map position: 2  
A:Introns: 27/3; 53/2; 138/2

Query Match	67.98	Score 38	DB 2	Length 207
Similarity	63.68	Pred. No. 23		
Best Local	7	Conservative	3	Mismatches
			1	Indels
				Gaps
				0

QY	2	IGICVAVTVAI	12
		11: :111:11	
Db	12	IGLVIATVIAI	22

RESULT 14

BH03705  
Phosphonates transport system (permease) BH0442 [imported] - *Bacillus halodurans* (strain  
C:Species: *Bacillus halodurans*  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: BH3705  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A:Reference number: A63650; MUID:20263314  
A:Accession: BH3705  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-267 <STO>  
A:Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BAB04161.1; GSPDB:GN00  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH0442  
C:Superfamily: phnE protein

Query Match	67.9%	Score 38	DB 2	Length 267
Best Local	Similarity 41.7%	Pred. No. 28		
Matches	5, Conservative	6, Mismatches	1, Indels	0, Gaps

```
QY      1 LGICVAVTVAI 12
          ::|||::: ::||
Db      84 IVGICLSIVLAI 95
```

## RESULT 15

D70729  
probable peptide transport system permease - Mycobacterium tuberculosis (strain H37Rv)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: D70729  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno  
 A:Reference number: A70500; M01D:98295987  
 A:Accession: D70729  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-291 <COL>  
 A:Cross-references: GB:577164; GB:AL123456; NID:g3261615; PIDN:CAB01002.1; PID:g32616  
 A:Experimental source: strain H37Rv  
 C:Genetics:  
 C:Gene: oppC  
 C:Superfamily: oligopeptide permease protein oppB

Query Match	67.9%	Score 38;	DB 2;	Length 291;
Best Local Similarity	87.5%	Pred. NO. 30;		
Matches 7;	Conservative 1;	Mismatches 0;	Gaps 0;	

QY 1 LIGICVAV 8  
 |||||  
 Db 86 LIGVCVAV 93

Search completed: November 13, 2001, 14:28:19  
Job time: 125 sec

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:34:24 ; Search time 17.9 seconds

(without alignments)  
22.965 Million cell updates/sec

Title: US-09-412-558-2

Sequence: 1 LIGICAVTVAI 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	100.0	68	1 VA13_VARY	P33838 variola vit
2	56	100.0	70	1 VA13_VACCC	P20990 vaccinia vi
3	42	75.0	569	1 ILIR_HUMAN	P14778 homo sapien
4	40	71.4	297	1 STX4_HUMAN	O12846 homo sapien
5	40	71.4	298	1 STX4_MOUSE	P70452 mus musculu
6	40	71.4	298	1 STX4_RAT	O08850 rattus norv
7	39	69.6	354	1 CARC_HUMAN	O43570 homo sapien
8	38	67.9	291	1 YC82_MYCTU	Q10623 mycobacteri
9	38	67.9	311	1 YCAY_CLOK1	P38943 Clostridium
10	38	67.9	469	1 SECY_SULSO	O9UX84 sulfolobus
11	37	66.1	141	1 HBA_BRATR	P14525 bradyopus tr
12	37	66.1	359	1 MRAY_MYCTU	O06221 mycobacteri
13	37	66.1	415	1 TNAB_ECOLI	P23173 escherichia
14	36	64.3	141	1 HBA1_ICO1G	P18974 iguana igua
15	36	64.3	176	1 CYB_EUMGL	Q34462 eumops glau
16	36	64.3	1472	1 ABC2_MOUSE	P41234 mus musculu
17	36	64.3	2233	1 RRPL_P13H4	P12577 human para
18	35	62.5	134	1 CYB_ANOCU	Q31664 anura caud
19	35	62.5	134	1 CYB_CHISL	Q34254 chiroderma
20	35	62.5	134	1 CYB_CHITR	Q34260 chiroderma
21	35	62.5	141	1 HBA_MIGA	P41331 microcephal
22	35	62.5	170	1 YR5_YEAST	P46991 saccharomyc
23	35	62.5	176	1 CYB_GLOS	Q33487 glossophaga
24	35	62.5	176	1 CYB_TADBR	Q35994 tadarida br
25	35	62.5	184	1 VA33_VARY	P33850 variola vit
26	35	62.5	209	1 CYB_RHIF	O21298 rhinolophus
27	35	62.5	291	1 Y4TO_RHISN	O53192 rhizobium s
28	35	62.5	457	1 YDHE_ECOLI	P37340 escherichia
29	35	62.5	986	1 EPA4_CHICK	O07496 gallus gall
30	35	62.5	3341	1 POLG_MGFA	P33515 genome po
31	34	60.7	130	1 CCKN_RANCA	P80344 rana caateso
32	34	60.7	134	1 CYB_PITSU	O78761 pitomys sub
33	34	60.7	245	1 ERB2_MOUSE	P70424 mus musculu

34	34	60.7	247	1 CD8A_MOUSE	P01731 mus musculu
35	34	60.7	406	1 U1A3_VZVU	P09273 varicella-z
36	34	60.7	540	1 KERB_AVIEU	P13723 avian eryth
37	34	60.7	604	1 KERB_AVIER	P00535 avian eryth
38	34	60.7	619	1 YOL8_CAEEL	Q02335 avian eryth
39	34	60.7	634	1 KERB_ALV	P00534 caenorhabdi
40	34	60.7	953	1 YMBG_YEAST	Q03516 avian leuko
41	34	60.7	1210	1 EGFR_HUMAN	P00533 homo sapien
42	34	60.7	1210	1 EGFR_MOUSE	O01279 mus musculu
43	34	60.7	1254	1 ERB2_MESAU	Q60553 mesocricetu
44	34	60.7	1255	1 ERB2_HUMAN	P04626 homo sapien
45	34	60.7	1257	1 ERB2_RAT	P06494 rattus norv

## ALIGNMENTS

```

RESULT 1
ID VA13_VARY STANDARD; PRT; 68 AA.
AC P33838;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
DE PROTEIN A13.
OS A13L OR A14L.
OS Variola virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10255;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INDIA-1967 / ISOLATE IND3;
RX MEDLINE=93202281; PubMed=8384129;
RA Sincelnikov S.N., Bilov V.M., Sandakhchiev L.S.;
RT "Genes of variola and vaccinia viruses necessary to overcome the host
RT protective mechanisms.";
RL FEBS Lett. 319:80-83(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BANGLADESH-1975;
RX MEDLINE=94088747; PubMed=8264798;
RA Massing R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
RA Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
RA Selivanov N.A., Cavaliaro K.F., Kerlavage A.R., Mahy B.W.J.,
RA Venter C.J.;
RT "Potential virulence determinants in terminal regions of variola
RT smallpox virus genome.";
RL Nature 366:748-751(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X69198; CAA49058.1; -;
DR EMBL; L22579; AAA60865.1; -;
DR PIR; E36849; E36849.
SQ SEQUENCE 68 AA; 7541 MW; 15BB909719033147 CRC64;

```

Query Match 100.0%; Score 56; DB 1; Length 68;

Best Local Similarity 100.0%; Pred. No. 0.0061;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LIGICAVTVAI 12  
DB 7 LIGICAVTVAI 18

```
RESULT 2
ID V133 VACC STANDARD: PRT: 70 AA.
AC P20990;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE PROTEIN A13.
GN A13.
OS Vaccinia virus (strain Copenhagen).
OC Viruses: dsDNA viruses, no RNA stage: Poxviridae; Chordopoxvirinae.
OC Orthopoxvirus
OC NCBI_Taxid=10249;
RN [1]
RP MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paolletti E.;
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M35027; AAA48135.1;
CC DR PIR: F42518; F42518; DB2F50B6E75F7955 CRC64;
CC SQ SEQUENCE 70 AA; 7696 MW; DB2F50B6E75F7955 CRC64;

Query Match 100.0%; Score 56; DB 1; Length 70;
Best Local Similarity 100.0%; Pred. No. 0.0063;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LGICVAVTVAI 12
   111111111111
Db 7 LGICVAVTVAI 18

RESULT 3
ID ILIR_HUMAN STANDARD: PRT: 569 AA.
AC P14778;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE INTERLEUKIN-1 RECEPTOR, TYPE I PRECURSOR (IL-1R-1) (IL-1R-ALPHA)
DE (P80) (ANTIGEN CD121A).
DE IL1R1 OR IL1RA OR IL1R.
GN Homo sapiens (human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_Taxid=9606;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RX MEDLINE=90096789; PubMed=2532321;
RA Chua A.O., Gubler U.;
RT "Sequence of the cDNA for the human fibroblast type interleukin-1
RT receptor.";
RL Nucleic Acids Res. 17:10114-10114(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
```

```
RX MEDLINE=90046906; PubMed=2530587;
RA Sims J.E., Acres R.B., Grubin C.E., McMahon C.J., Wignall J.M.,
RA March C.J., Dower S.K.;
RT "Cloning the interleukin 1 receptor from human T cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:8946-8950(1989).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 18-332 IN COMPLEX WITH IL1B.
RX MEDLINE=97215903; PubMed=9062193;
RA Vigers G.P., Anderson L.J., Caffes P., Brandhuber B.J.;
RT "Crystal structure of the type I interleukin-1 receptor complexed
RT with interleukin-1beta.";
RL Nature 386:190-194(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 21-331 IN COMPLEX WITH IL1RA.
RX MEDLINE=97215904; PubMed=9062194;
RA Schreuder H., Tardif C., Trump-Kallmeyer S., Soffientini A.,
RA Sarubbi E., Akesson A., Bowlin T., Yanofsky S., Barrett R.W.;
RT "A new cytokine-receptor binding mode revealed by the crystal
RT structure of the IL-1 receptor with an antagonist.";
RL Nature 386:194-200(1997).
CC -1- FUNCTION: RECEPTOR FOR INTERLEUKIN-1 ALPHA (IL-1A), BETA (IL-1B),
CC AND INTERLEUKIN-1 RECEPTOR ANTAGONIST PROTEIN (IL-1RA). BINDING TO
CC THE AGONIST LEADS TO THE ACTIVATION OF NF-KAPPA B.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD121a entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd121a.htm"
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16896; CAA34773.1;
CC DR EMBL: M27492; AAA59137.1;
CC DR PIR: S06928; S06928.
CC DR PIR: A36187; A36187.
CC DR PDB: 1YTB; 04-FEB-98.
CC DR PDB: 1IRA; 17-JUN-98.
CC DR MIM: 147810;
CC DR InterPro: IPR000157;
CC DR InterPro: IPR003006;
CC DR Pfam: PF01582; TIR: 1.
CC DR Pfam: PF00047; 19; 2.
CC KW Immunoglobulin domain; Transmembrane; Glycoprotein; Receptor; Signal;
CC 3D-structure.
CC FT SIGNAL 1 17
CC FT CHAIN 18 569
CC FT DOMAIN 18 336
CC FT TRANSMEM 337 356
CC FT DOMAIN 357 569
CC FT DOMAIN 37 103
CC FT DOMAIN 135 203
CC FT DOMAIN 241 319
CC FT DISULFID 23 104
CC FT DISULFID 44 96
CC FT DISULFID 121 164
CC FT DISULFID 142 196
CC FT DISULFID 248 312
CC FT CARBOHYD 100 100
CC FT CARBOHYD 193 193
CC FT CARBOHYD 233 233
CC FT CARBOHYD 249 249
CC FT CARBOHYD 263 263
CC FT CARBOHYD 297 297
CC SQ SEQUENCE 569 AA; 65402 MW; 5BAAB3F8F0225C25 CRC64;

Query Match 75.0%; Score 42; DB 1; Length 569;
Best Local Similarity 66.7%; Pred. No. 6; 6;
```

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
:||||:| |  
Db 338 MLCVTVTVII 349

RESULT 4  
STX4\_HUMAN STANDARD; PRT; 297 AA.  
AC Q12846; Q15525;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE SYNAXIN 4.  
GN STX4 OR STX4.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA MEDLINE=9426173; PubMed=8206394;  
RA LI H., Hodge D.R., Pel G.K., Seth A.;  
RT Isolation and sequence analysis of the human syntaxin-encoding  
RT gene";  
RL Gene 143:303-304(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skeletal muscle;  
RA MEDLINE=96332494; PubMed=8760387;  
RA Jagadeish M.N., Fernandez C.S., Hewish D.R., Macaulay S.L.,  
RA Gough K.H., Griscov J., Verkuylen A., Cosgrove L., Alafaci A.,  
RA Fremel M.D., Ward C.W.;  
RT Insulin-responsive tissues contain the core complex protein SNAP-25  
RT (syntaxin-4 and synaptobrevin 1 and 2.");  
RT Biochem. J. 317:945-954(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Peripheral blood neutrophils;  
RA Nabokina S., Lazo P.A., Mollinedo F.;  
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Lens epithelium;  
RA Rae J.L., Shepard A.R.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: POTENTIALLY INVOLVED IN DOCKING OF SYNAPTIC VESICLES  
AT PRESYNAPTIC ACTIVE ZONES.  
CC -1- SIMILARITY: BELONGS TO THE SYNAXIN/EPIMORPHIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U07158; AAA20967.1; -  
CC EMBL: X85784; CAA59769.1; -  
CC EMBL: AJ000541; CAA04174.1; -  
CC EMBL: AF026007; AAB8810.1; -  
CC HSSP: P32851; IBR0.  
CC MIM: 186591; -  
CC InterPro: IPR000017; -  
CC Pfam: PF00804; SYNAXIN; 1.  
CC PROSITE: PS00914; SYNAXIN; 1.  
CC Neurotransmitter transport; Coiled coil; Transmembrane.  
FT DOMAIN 43 163 COILED COIL (POTENTIAL).  
FT DOMAIN 199 222 COILED COIL (POTENTIAL).  
FT

FT TRANSMEM 276 296 POTENTIAL.  
FT CONFLICT 174 174 E -> D (IN REF. 1).  
FT CONFLICT 269 269 A -> V (IN REF. 1).  
SO SEQUENCE 297 AA; 34180 MW; 5084FDIC49A868AA CRC64;

Query Match 71.4%; Score 40; DB 1; Length 297;  
Best Local Similarity 58.3%; Pred. No. 8.3;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
:||||:| |  
Db 275 LIAICVSTVVL 286

RESULT 5  
STX4\_MOUSE STANDARD; PRT; 298 AA.  
ID STX4\_MOUSE  
AC P70452;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE SYNAXIN 4.  
GN STX4 OR STX4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=97197781; PubMed=9045631;  
RA Teliam J.T., Macaulay S.L., McIntosh S., Hewish D.R., Ward C.W.,  
RA James D.E.;  
RT Characterization of Munc-18c and syntaxin-4 in 3T3-L1 adipocytes.  
RT Putative role in insulin-dependent movement of GLUT-4.";  
RL J. Biol. Chem. 272:6179-6186(1997).  
CC -1- FUNCTION: POTENTIALLY INVOLVED IN DOCKING OF SYNAPTIC VESICLES  
AT PRESYNAPTIC ACTIVE ZONES.  
CC -1- SIMILARITY: BELONGS TO THE SYNAXIN/EPIMORPHIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U76832; AAB18991.1; -  
CC HSSP: P32851; IBR0.  
CC MGD: MGI:893577; STX4.  
CC InterPro: IPR000017; -  
CC Pfam: PF00804; SYNAXIN; 1.  
CC PROSITE: PS00914; SYNAXIN; 1.  
CC Neurotransmitter transport; Coiled coil; Transmembrane.  
FT DOMAIN 38 163 COILED COIL (POTENTIAL).  
FT DOMAIN 199 223 COILED COIL (POTENTIAL).  
FT TRANSMEM 275 295 POTENTIAL.  
SO SEQUENCE 298 AA; 34165 MW; FCD1477E1126CEC1 CRC64;

Query Match 71.4%; Score 40; DB 1; Length 298;  
Best Local Similarity 66.7%; Pred. No. 8.4;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVAI 12  
:||||:| |  
Db 275 MIAICVSTVVL 286

RESULT 6  
STX4\_RAT STANDARD; PRT; 298 AA.  
ID STX4\_RAT

```

AC 008850;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SYNAXIN 4.
GN SYNAXIN 4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP MEDLINE=9336759; PubMed=7690687;
RA Bennett M.K., Hazuka C.D., Scheller R.H.,
RA Fleming A.M., Hazuka C.D., Scheller R.H.,
RT "The syntaxin family of vesicular transport receptors."
RL Cell 74:863-873(1993).
CC -1- FUNCTION: POTENTIALLY INVOLVED IN DOCKING OF SYNAPTIC VESICLES
CC -1- TISSUE SPECIFICITY: HEART, SPLEEN, SKELETAL MUSCLE, AND KIDNEY.
CC -1- SIMILARITY: BELONGS TO THE SYNAXIN/EPIMORPHIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L20821; AAA03046.1; -
DR HSSP; P32851; 1BR0.
DR InterPro: IPR000017; -
DR Pfam; PF00804; Syntaxin: 1.
DR PROSITE; PS00914; SYNTAXIN: 1.
DR Neurotransmitter transport; Coiled coil; Transmembrane.
KW DOMAIN 38 163 COILED COIL (POTENTIAL).
FT 199 223 COILED COIL (POTENTIAL).
FT DOMAIN 275 295 POTENTIAL.
FT TRANSMEM 298 AA; 34209 MW; 9EB54270DFB3CB96 CRC64;
FT SEQUENCE
SO
Query Match 71.4%; Score 40; DB 1; Length 288;
Best Local Similarity 66.7%; Pred. NO. 8.4;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 1 ILIGICVAVTVAI 12
DB 275 MIAICVSVTVLI 286

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 95:7608-7613(1998).
RN [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=98445416; PubMed=9770531;
RX Ivanov S.V., Kuzmin I., Wei M.-H., Pack S., Gell L., Johnson B.E.,
RA Standridge E.J., Lerman M.I.;
RT Down-regulation of transmembrane carbonic anhydrases in renal cell
RT carcinoma cell lines by wild-type von Hippel-Lindau transgene.
RL Proc. Natl. Acad. Sci. U.S.A. 95:12596-12601(1998).
CC -1- FUNCTION: REVERSIBLE HYDRATATION OF CARBON DIOXIDE.
CC -1- CATALYTIC ACTIVITY: H(2)CO(3) = CO(2) + H(2)O.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PANCREAS, OVARY, AND TESTIS.
CC -1- PROSTATE AND MODERATELY EXPRESSED IN PANCREAS, OVARY, AND TESTIS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC-TYPE CARBONIC ANHYDRASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF051882; AAC39789.1; -
DR EMBL; AF037335; AAC63952.1; -
DR MIM; 603263; -
DR InterPro: IPR001148; -
DR Pfam; PF00194; carb.anhydrase: 1.
DR PROSITE; PS00162; EUK CO2 ANHYDRASE: 1.
KW Lyase; Zinc; Transmembrane; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 354 CARBONIC ANHYDRASE XII.
FT DOMAIN 25 301 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 302 322 POTENTIAL.
FT DOMAIN 323 354 CYTOPLASMIC (POTENTIAL).
FT METAL 119 119 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 121 121 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 145 145 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 354 AA; 39451 MW; 9016216BF2CA6C0C CRC64;
SO
Query Match 69.6%; Score 39; DB 1; Length 354;
Best Local Similarity 50.0%; Pred. NO. 14;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
OY 1 ILIGICVAVTVAI 12
DB 313 ILIGICVAVTVSI 324

```

```

RESULT 8
ID YC82_MYCTU STANDARD; PRT; 291 AA.
AC 010623; Q50698;
DT 01-NOV-1997 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE PEPTIDE TRANSPORT PERMEASE PROTEIN RV1282C.
DE RV1282C OR MTCY373.01C OR MTCY383.01.
GN Mycobacterium tuberculosis.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Bacteria; Firmicutes; Actinobacteriae; Mycobacteriaceae; Mycobacterium.
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
overexpressed in some renal cell cancers."

```

RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Fellmuth T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jørgensen K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osbourne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellon S., Squares S., Squires R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
 CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-  
 CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPPBC  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z77164; CAB01002.1; -  
 CC DR Tuberculin: Rv1282c; -  
 CC DR InterPro: IPR000515; -  
 CC DR Pfam: PF00528; BPD\_TRANSP; 1.  
 CC DR PROSITE: PS00402; BPD\_TRANSP\_INN\_MEMBER; 1.  
 CC KW Hypothetical protein; Transport; Peptide transport; Transmembrane.  
 CC FT TRANSMEM 22 42 POTENTIAL.  
 CC FT TRANSMEM 85 105 POTENTIAL.  
 CC FT TRANSMEM 116 136 POTENTIAL.  
 CC FT TRANSMEM 142 162 POTENTIAL.  
 CC FT TRANSMEM 209 229 POTENTIAL.  
 CC FT TRANSMEM 247 267 POTENTIAL.  
 CC SO SEQUENCE 291 AA; 31374 MW; F6BB51906154CF9F CRC64;  
 QY 1 LIGICVAV 8  
 DB 86 LIGICVAV 93  
 RESULT 9  
 YCAY\_CLOKL STANDARD; PRT; 311 AA.  
 AC P38943;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HYPOTHELICAL 33.5 KDA PROTEIN IN CAT1 5'REGION (ORF).  
 OS Clostridium kluyveri.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
 OC Clostridium;  
 OC MCB1\_TaxID=1534;  
 OX NCBI\_TaxID=11;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 555;  
 RX MEDLINE=96146540; PubMed=8550525;  
 RA Soehling B., Gottschalk G.;  
 RT "Molecular analysis of the anaerobic succinate degradation pathway in  
 RT Clostridium kluyveri.";  
 RL J. Bacteriol. 178:871-880(1996).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -1- SIMILARITY: BELONGS TO THE UPF0077 (PECM) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L21902; AAA92345.1; -  
 CC DR InterPro: IPR000620; -  
 CC DR Pfam: PF00892; DUF6; 2.  
 CC KW Hypothetical protein; Transmembrane.  
 CC FT TRANSMEM 6 26 POTENTIAL.  
 CC FT TRANSMEM 33 53 POTENTIAL.  
 CC FT TRANSMEM 70 90 POTENTIAL.  
 CC FT TRANSMEM 97 117 POTENTIAL.  
 CC FT TRANSMEM 123 143 POTENTIAL.  
 CC FT TRANSMEM 155 175 POTENTIAL.  
 CC FT TRANSMEM 185 205 POTENTIAL.  
 CC FT TRANSMEM 219 239 POTENTIAL.  
 CC FT TRANSMEM 244 264 POTENTIAL.  
 CC FT TRANSMEM 265 285 POTENTIAL.  
 CC SO SEQUENCE 311 AA; 33507 MW; A5E7CB9E9C9ADF0B CRC64;  
 QY 1 LIGICVAVTVAI 12  
 DB 155 LIGICFALVAV 166  
 RESULT 10  
 SECY\_SULSO STANDARD; PRT; 469 AA.  
 AC Q9UX84;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PREPROTEIN TRANSLOCASE SECY SUBUNIT.  
 GN SECY OR C10.035.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
 OC NCBI\_TaxID=2287;  
 OX NCBI\_TaxID=11;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1617 / P2;  
 RX MEDLINE=20165948; PubMed=10701121;  
 RA Charlebois R.L., Singh R.K., Chan-Weber C.C.-Y., Allard G., Chow C.,  
 RA Confalonieri F., Curtis B., Duguet M., Erasus G., Faguy D.,  
 RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,  
 RA Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.L., She O., F.,  
 RA St Jean A., van der Oost J., Young F., Zivanovic Y., Doolittle W.F.,  
 RA Ragan M.A., Sensen C.W.;  
 RT "Gene content and organization of a 281-kbp contig from the genome of  
 RT the extremely thermophilic archaeon, Sulfolobus solfataricus p2.";  
 RL Genome 43:116-136(2000).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SRCA AND SECY  
 CC TO ALLOW THE TRANSLLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE,  
 CC BY FORMING PART OF A CHANNEL.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE SECY/SEC61-ALPHA FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y18930; CAB57608.1; -  
 CC DR InterPro: IPR002208; -  
 CC DR PRINTS: PR00303; SECYRNLCASE.

Query Match	66.18;	Score 37;	DB 1;	Length 141;
Best Local Similarity	58.38;	Pred. No. 14;		
Matches	7;	Conservative	3;	Mismatches
			2;	Indels
				0;
Gaps				0;
QY	1	LIGICVAVYVAI	12	
Db	100	LGHCVCVTLAL	111	

MRAY	MYCTU	STANDARD:	PRT:	359 AA.
AC	006221:			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, last sequence update)			
DT	30-MAY-2000 (Rel. 39, last annotation update)			
DE	PHOSPHO-N-ACETYLUMURAMOYL-PENTAPEPTIDE-TRANSFERASE (EC 2.7.8.13)			
DE	(UDP-MURNAc-PENTAPEPTIDE PHOSPHOTRANSFERASE).			
DN	MRAY OR MURX OR RV2156C OR MTC270.12.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria: Firmicutes: Actinobacteria; Actinobacteridae;			
OC	Actinomycetales: Corynebacterineae: Mycobacteriaceae: Mycobacterium.			
ON	NCBI_TaxID=1773;			
RM	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=96295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekaia F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsbury T., Jagers K., Feltwell T., Gentle S., Hamlin N., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Ralstead M.A., Rogers J.,			
RA	Rafter S., Seeger K., Skelton S., Squires S., Squires R., Sulston J.E.,			
RA	Taylor K., Whitehead S., Barrell B.G.,			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RL	complete genome sequence."			
RL	Nature 393:537-544(1998).			
CC	-1- FUNCTION: FIRST STEP OF THE LIPID CYCLE REACTIONS IN THE			
CC	BIOSYNTHESIS OF THE CELL WALL PEPTIDOGLYCAN (BY SIMILARITY).			
CC	-1- CATALYTIC ACTIVITY: UDP-N-ACETYLMURAMOYL-L-ALANYL-D-GLUTAMYL-L-			
CC	LYSYL-D-ALANYL-D-ALANINE + UNDECAPRENYL PHOSPHATE = UDP +			
CC	N-ACETYLMURAMOYL-L-ALANYL-D-GLUTAMYL-L-LYSYL-D-ALANYL-D-ALANINE-			
CC	DIPHOSPHOUNDECAPRENOL.			
CC	-1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.			
CC	-1- SUBCELLULAR LOCATION: INTERIAL MEMBRANE PROTEIN (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 4. MRAY			
CC	SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: Z95388; CAB08671.1; -			
DR	Tuberculist: RV2156C; -			
DR	InterPro: IPR001715; -			
DR	Pfam: PF00963; GLYCOS_transf_4; 1.			
DR	PROSITE: PS01347; MRAY_1; 1.			
DR	PROSITE: PS01346; MRAY_2; 1.			
KW	peptidoglycan synthesis; Cell division; Transferase; Transmembrane.			
FT	TRANSMEM 3 23			
FT	TRANSMEM 55 75			
FT	TRANSMEM 80 100			
FT	TRANSMEM 117 137			
FT	TRANSMEM 156 176			
FT	TRANSMEM 187 207			
FT	TRANSMEM 231 251			
FT	TRANSMEM 255 275			
FT	TRANSMEM 280 300			
FT	TRANSMEM 334 354			
FT	SEQUENCE 359 AA; 37713 MW; 27E979BC77BA28C6 CRC64;			

Db 5 LIAVAVATVSI 16

RESULT 13

TNAB\_ECOLI STANDARD; PRT; 415 AA.

ID TNAB\_ECOLI

AC P23173;

DT 01-NOV-1991 (Rel. 20, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE LOW AFFINITY TRYPTOPHAN PERMEASE.

GN TNAB OR TRPP.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RC MEDLINE-91216998; PubMed-2022620;

RA Sarero J.P., Mooney P.J., Golnick P., Yanofsky C., Pittard A.J.;

RT "A new family of integral membrane proteins involved in transport of aromatic amino acids in Escherichia coli.";

RL J. Bacteriol. 173:3231-3234(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RC MEDLINE-93315143; PubMed-7686882;

RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;

RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication.";

RL Genomics 16:551-561(1993).

RN [3]

RP SEQUENCE OF 1-13 FROM N.A.

RC STRAIN-K12;

RC MEDLINE-82007678; PubMed-6268608;

RA Deeley M.C., Yanofsky C.;

RT "Nucleotide sequence of the structural gene for tryptophanase of Escherichia coli K-12.";

RL J. Bacteriol. 147:787-796(1981)

CC -1- FUNCTION: INVOLVED IN THE TRYPTOPHAN TRANSPORT ACROSS THE CYTOPLASMIC MEMBRANE. PLAY A ROLE IN TRANSPORTING TRYPTOPHAN WHICH IS TO BE USED CATABOLICALLY.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.

CC -1- INDUCTION: BY TRYPTOPHAN. IS SUBJECT TO CATABOLIC REPRESSION.

CC -1- SIMILARITY: BELONGS TO THE MTR / TNAB / TYRO PERMEASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: K00032; AAA24677.1; -

DR EMBL: M59914; AAA62792.1; -

DR EMBL: L10328; AAA62060.1; -

DR EMBL: AE000448; AAC76732.1; -

DR PIR: A39412; A39412.

DR EcoGene: EGI1006; tnab.

DR InterPro: IPR002091; -

DR PRINTS: PR00166; AROAARMEASE.

DR PROSITE: PS00594; AROMATIC\_AA\_PERMEASE\_1; 1.

KM Transport; amino-acid transport; Transmembrane; Inner membrane.

FT TRANSMEM 12 32 POTENTIAL.

FT TRANSMEM 34 54 POTENTIAL.

FT TRANSMEM 87 107 POTENTIAL.

FT TRANSMEM 128 148 POTENTIAL.

FT TRANSMEM 154 174 POTENTIAL.

FT TRANSMEM 192 212 POTENTIAL.

FT TRANSMEM 230 250 POTENTIAL.

FT TRANSMEM 287 307 POTENTIAL.

FT TRANSMEM 327 347 POTENTIAL.

FT TRANSMEM 349 369 POTENTIAL.

FT TRANSMEM 388 408 POTENTIAL.

FT CONFLICT 3 3 D -> V (IN REF. 3).

SQ SEQUENCE 415 AA; 45210 MW; 1E34C7DE3DE69D9 CRC64;

Query Match 66.1%; Score 37; DB 1; Length 415;

Best Local Similarity 54.5%; Pred. No. 34;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVA 11

Db 128 IGVGICTAIFVA 138

RESULT 14

HBAL\_IGUG STANDARD; PRT; 141 AA.

ID HBAL\_IGUG

AC P18974;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE HEMOGLOBIN ALPHA-1 CHAIN.

OS Iguana iguana (Common Iguana).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Lepidosauria; Squamata; Iguania; Iguanidae; Iguaninae; Iguana.

OX NCBI\_TaxID=8517;

RN [1]

RP SEQUENCE.

RA Ruecknagel K.P., Braunitzer G., Wiesner H.;

RT "Hemoglobins of reptiles. The primary structures of the alpha I- and beta I-chains of common Iguana (Iguana iguana) hemoglobin.";

RL Biol. Chem. Hoppe-Seyler 369:1143-1150(1988).

DR PIR: S01664; HAIG1.

DR HSSP: P01922; 1FDD.

DR InterPro: IPR000971; -

DR Pfam: PF00042; globin; 1.

DR PROSITE: PS01033; GLOBIN; 1.

KM Heme; Oxygen transport; Respiratory protein; Erythrocyte.

FT METAL 58 58 IRON (HEME DISTAL LIGAND).

FT METAL 87 87 IRON (HEME PROXIMAL LIGAND).

SQ SEQUENCE 141 AA; 15439 MW; 9B56DCCCE5DCBF97 CRC64;

Query Match 64.3%; Score 36; DB 1; Length 141;

Best Local Similarity 54.5%; Pred. No. 21;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVAVTVA 11

Db 100 LIGHCILVTVA 110

RESULT 15

CYB\_EUMGL STANDARD; PRT; 176 AA.

ID CYB\_EUMGL

AC Q34462;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE CYTOCHROME B (FRAGMENT).

GN MTCYB OR COB OR CYTB.

OS Eumops glaucinus.

OC Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Chiroptera; Microchiroptera; Molossidae; Eumops.

OX NCBI\_TaxID=27619;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-LSUM2 27212; TISSUE-liver, and kidney;

RA Sudman P.D., Barkley L.J., Hafner M.S.;

```

RT "Familial affinity of Tomopeas ravus (Chiroptera) based on protein
RL electrophoretic and cytochrome b sequence data.";
CC J. Mammal. 75:365-377(1994).
CC -1- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -1- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L19719; AAA17765.1; -.
CC InterPro; IPR000179; -.
CC DR Pfam; PF00033; cytochrome_b_N; 1.
CC DR PROSITE; PS00192; CYTOCHROME_B_HEME; 1.
CC DR PROSITE; PS00193; CYTOCHROME_B_OO; PARTIAL.
CC KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
CC Heme.
CC FT METAL 83 83 IRON 1 (HEME B562 AXIAL LIGAND).
CC FT METAL 97 97 IRON 2 (HEME B566 AXIAL LIGAND).
CC FT NON_TER 176 176
CC SQ SEQUENCE 176 AA; 19583 MW; 6C948ACA8905F1A9 CRC64;

Query Match 64.3%; Score 36; DB 1; Length 176;
Best Local Similarity 60.0%; Pred. No. 25;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 LIGICVAVTV 10
   1:|11:11:
DB 36 LIGICLAVOI 45

```

Search completed: November 13, 2001, 14:34:25  
 Job time: 396 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:34:00 ; Search time 46.47 Seconds  
(without alignments)  
34.165 Million cell updates/sec

Title: US-09-412-558-2

Perfect score: 56

Sequence: 1 LIGICVAVTVAI 12

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

SPREMBL\_16:\*

1: sp.archaea:\*

2: sp.bacteria:\*

3: sp.fungi:\*

4: sp.human:\*

5: sp.invertebrate:\*

6: sp.mammal:\*

7: sp.mhc:\*

8: sp.organelle:\*

9: sp.phage:\*

10: sp.plant:\*

11: sp.potent:\*

12: sp.unclassified:\*

13: sp.vertebrate:\*

14: sp.virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	100.0	70	14	057225 Vaccinia vi
2	56	100.0	70	14	09JF84 Vaccinia vi
3	52	92.9	68	14	089163 variola vir
4	40	71.4	237	2	025593
5	39	69.6	470	3	09P8G5
6	38.5	68.8	573	5	017654
7	38	67.9	87	14	09E207
8	38	67.9	114	2	087210
9	38	67.9	207	5	017160
10	38	67.9	257	2	09KFN7
11	38	67.9	353	2	062368
12	37	66.1	135	2	0914D5
13	37	66.1	166	2	09RF47
14	37	66.1	193	2	045385
15	37	66.1	218	2	031603
16	37	66.1	272	5	09M4U8
17	37	66.1	380	8	09T6R5
18	37	66.1	434	2	087996
19	37	66.1	436	10	09S374

20	37	66.1	502	10	03SP14	09SP14 alonsoa mer
21	37	66.1	510	10	09SP15	09SP15 asarina bar
22	37	66.1	552	5	P91765	P91765 myzus persi
23	37	66.1	599	2	P71787	P71787 mycobacteri
24	37	66.1	944	4	014697	014697 homo sapien
25	37	66.1	966	4	09P0X0	09P0X0 homo sapien
26	37	66.1	966	11	008794	008794 mus musculu
27	37	66.1	973	3	09V707	09V707 schizosacch
28	37	66.1	2233	14	081890	081890 human para
29	37	66.1	2233	14	09J7D6	09J7D6 bovine para
30	37	66.1	2233	14	09J7D0	09J7D0 bovine para
31	37	64.3	111	2	025112	025112 helicobacte
32	36	64.3	134	8	09XMC9	09XMC9 morioops me
33	36	64.3	196	2	09Z663	09Z663 zymomonas m
34	36	64.3	231	5	09TXJ7	09TXJ7 caenorhabd
35	36	64.3	266	5	0917L7	0917L7 drosophila
36	36	64.3	296	10	09XEX3	09XEX3 zea mays (m
37	36	64.3	350	7	031062	031062 sus scrofa
38	36	64.3	364	10	09LM68	09LM68 arabidopsis
39	36	64.3	366	5	09XWT7	09XWT7 caenorhabd
40	36	64.3	379	8	09GAN1	09GAN1 carollia pe
41	36	64.3	380	8	09ZEC9	09ZEC9 upupa epops
42	36	64.3	422	10	040543	040543 nicotiana t
43	36	64.3	440	4	075655	075655 homo sapien
44	36	64.3	453	4	09NRX5	09NRX5 homo sapien
45	36	64.3	453	11	09QZ18	09QZ18 mus musculu

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	70 AA.
057225				
ID	057225			
AC	057225			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	STRUCTURAL PROTEIN P8.			
GN	MYA124L.			
OS	Vaccinia virus (strain Ankara).			
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;			
OC	Orthopoxvirus.			
OX	NCBI_TaxID=126794;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ANKARA;			
RA	Antoine G., Scheiflinger F., Falkner F.G., Dörner F.;			
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: U94848; AAB96464.1; -			
SO	SEQUENCE 70 AA; 7644 MW; C5E59815B7BF7955 CRC64;			
Query Match	100.0%; Score 56; DB 14; Length 70;			
Best Local Similarity	100.0%; Pred. No. 0.015;			
Matches 12; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 LIGICVAVTVAI 12			
DB	7 LIGICVAVTVAI 18			
RESULT	2			
ID	09JF84	PRELIMINARY;	PRT;	70 AA.
AC	09JF84;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)			
DE	TALAI.			
OS	Vaccinia virus (strain Tian Tan).			
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;			
OC	Orthopoxvirus.			

```

OX NCBL_TaxID=10253;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=TIAN TAN;
RA Jin O., Hou Y.D., Cheng N.H., Yao E.M., Cheng S.X., Yang X.K.,
RA Jing D.Y., Yu W.H., Yuan J.S., Ma X.J.;
RA "Complete genomic sequence of vaccinia virus (Tian Tan strain).";
RA Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF095689; AAF34007.1; -
SQ SEQUENCE 70 AA; 7668 MW; C43F50B6E75F7955 CRC64;

Query Match
Best Local Similarity 100.0%; Score 56; DB 14; Length 70;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGICVAVTVAI 12
Db 7 LGICVAVTVAI 18

RESULT 3
089163 PRELIMINARY; PRT; 68 AA.
ID 089163;
AC 089163;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE ORF4L.
GN A14L.
OS Variola virus, and variola minor virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBL_TaxID=10255, 53258;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Variola virus; STRAIN=GARCIA-1966;
RA Shchelkunov S.N., Totmenin A.V., Sosnovtsev S.V., Safronov P.F.,
RA Resenchuk S.M., Blinov V.M., Sandakhchiev L.S.;
RA Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Variola minor virus; STRAIN=GARCIA-1966;
RA Shchelkunov S.N., Totmenin A.V., Gutorov V.V., Safronov P.F.,
RA Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
RA Esposto J.J., Sosnovtsev S.;
RA "Analysis of the complete coding sequence of DNA of alastrim variola
RA minor virus strain Garcia-1966.";
RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: X76268; CAA53886.1; -
DR EMBL: Y16780; CAB54717.1; -
SQ SEQUENCE 68 AA; 7569 MW; 32BB87801903315E CRC64;

Query Match
Best Local Similarity 92.9%; Score 52; DB 14; Length 68;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LGICVAVTVAI 12
Db 7 LGICVAVTVAI 18

RESULT 4
025593 PRELIMINARY; PRT; 237 AA.
ID 025593;
AC 025593;
DT 01-JUN-1998 (TRENBLrel. 05, Created)
DT 01-JUN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE AMINO ACID ABC TRANSPORTER, PERMEASE PROTEIN (YCKU).
GN HP0939.
OS Helicobacter pylori (Campylobacter pylori).
```

```

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter
OX NCBL_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Klenk H.-P., Gill S., Dougherty B.A.,
RA Fleischmann R.D., Ketchum K.A., Adon7984.1; -
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khaliq H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Melman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL: AE000603; AAD07984.1; -
DR TIGR: HP0939; -
DR InterPro: IPR000515; -
DR Pfam: PF00528; BPD.transp. 1.
RW Hypothetical protein.
SQ SEQUENCE 237 AA; 26351 MW; EB3FA49863394A27 CRC64;

Query Match
Best Local Similarity 71.4%; Score 40; DB 2; Length 237;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 LGICVAVTVAI 12
Db 48 ILGICIAVFVAL 59

RESULT 5
09P8G5 PRELIMINARY; PRT; 470 AA.
ID 09P8G5;
AC 09P8G5;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE PUTATIVE SNARE PROTEIN.
GN SYN.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBL_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Gupta G.D., Heath B.I.;
RT "Neurospora crassa putative SNARE, similar to yeast SSO1P."
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF262964; AAF72704.1; -
DR InterPro: IPR00017; -
DR InterPro: IPR000727; -
DR Pfam: PF00804; Syntaxin; 1.
SQ SEQUENCE 470 AA; 52175 MW; 2FCBACFBCC60F934 CRC64;

Query Match
Best Local Similarity 69.6%; Score 39; DB 3; Length 470;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 IGICVAVTVAI 12
Db 295 LGICVAVTVAI 305

RESULT 6
017654 PRELIMINARY; PRT; 573 AA.
ID 017654
```

AC 017654;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)  
 DE C4166.2 PROTEIN.  
 GN C4166.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cummings P.;  
 RL Submitted (OCr-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Shownkeen R.,  
 RA Smaldon N., Smith A., Sonnhamer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";  
 RL Nature 368:32-38(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Lloyd C.R.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=99069613; PubMed=9851916;  
 none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 investigating biology.";  
 RL Science 283:2012-2018(1998).  
 DR EMBL; 261047; CAB02826.1; -;  
 DR EMBL; 283116; CAB60294.1; -;  
 DR EMBL; 281047; CAB60294.1; JOINED.  
 SQ SEQUENCE 573 AA; 66076 MW; 096FF19E591851A8 CRC64;

Query Match 68.8%; Score 38.5; DB 5; Length 573;  
 Best Local Similarity 61.5%; Pred. No. 88;  
 Matches 8; Conservative 4; Mismatches 0; Indels 1; Gaps 1;

OY 1 LIGICVA-VTVAI 12  
 DB 76 MIIICIAIDLTIAI 88

RESULT 7  
 09E207  
 AC 09E207; PRELIMINARY; PRT; 87 AA.  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE MEMBRANE PROTEIN.  
 OS Cercopithecine herpesvirus 7.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicelloviruses.  
 OX NCBI\_TaxID=35245;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.;  
 RT "Complete Sequence of the Simian Varicella Virus Genome.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF275348; AAG27247.1;  
 SQ SEQUENCE 87 AA; 9693 MW; 27491B2C589B63D2 CRC64;

Query Match 67.9%; Score 38; DB 14; Length 87;  
 Best Local Similarity 66.7%; Pred. No. 21;  
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 LIGICVA-VTV 9  
 DB 59 LVGCVAVLT 67

RESULT 8  
 087210  
 ID 087210; PRELIMINARY; PRT; 114 AA.  
 AC 087210;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE TRSB PROTEIN (TRAB).  
 OS Lactococcus lactis.  
 OG Plasmid pMRC01.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Lactococcus.  
 OX NCBI\_TaxID=1358;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-DPC3147;  
 RC MEDLINE=99000510; PubMed=9767571;  
 RA Dougherty B.A., Hill C., Weidman J.F., Richardson D.R., Venter J.C.,  
 RA Ross R.P.;  
 RT "Sequence and analysis of the 60 kb conjugative, bacteriocin-producing  
 plasmid pMRC01 from Lactococcus lactis DPC3147.";  
 RL Mol. Microbiol. 29:1029-1038(1998).  
 DR EMBL; AE001272; AAC55994.1; -;  
 KW Plasmid.  
 SQ SEQUENCE 114 AA; 11902 MW; 4374310433AF9669 CRC64;

Query Match 67.9%; Score 38; DB 2; Length 114;  
 Best Local Similarity 50.0%; Pred. No. 26;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 LIGICVA-VTVAI 12  
 DB 59 LVGCVAVLTFTI 70

RESULT 9  
 017160  
 ID 017160; PRELIMINARY; PRT; 207 AA.  
 AC 017160;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE C24H12.8 PROTEIN.  
 GN C24H12.8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RC MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Ritken L., Roopra A., Saunders D., Shownkeen R.,

RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sprout J., Wohlman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Rohlfing T., Wohlmann P., Biewald T.;  
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF025451; AAB71204.1; -;  
 DR InterPro: IPR001810; -;  
 DR Pfam: PF00646; F-box; 1;  
 DR PROSITE: PS50161; F-box; 1;  
 SO SEQUENCE 207 AA; 24415 MW; 4CB3C56F2F4FB8C CRC64;

Query Match 67.9%; Score 38; DB 5; Length 207;  
 Best Local Similarity 63.6%; Pred. No. 44;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 IGICVAVTVAI 12  
 DB 12 IGLVIAVTVAI 22

RESULT 10  
 ID Q9KFN7 PRELIMINARY; PRT; 267 AA.  
 AC Q9KFN7; 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE PHOSPHONATES TRANSPORT SYSTEM (PERMEASE).  
 GN BH0442.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=86655;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-C-125 / JCM 9153;  
 RA Takami H., Nakasone K., Takaki Y.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF001508; BAB04161.1; -;  
 DR InterPro: IPR000515; -;  
 DR Pfam: PF00528; BPD\_transp; 1;  
 SO SEQUENCE 267 AA; 28649 MW; 0E6D98B66428D1E CRC64;

Query Match 67.9%; Score 38; DB 2; Length 267;  
 Best Local Similarity 41.7%; Pred. No. 55;  
 Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
 DB 84 IGVICISIVLAI 95

RESULT 11  
 ID 062368 PRELIMINARY; PRT; 353 AA.  
 AC 062368;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)  
 DE T06G6.2 PROTEIN.

GN T06G6.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RA Kershaw J.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sprout J., Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 DR EMBL: Z81587; CAB04702.1; -;  
 SO SEQUENCE 353 AA; 40782 MW; FA75598667415E45 CRC64;

Query Match 67.9%; Score 38; DB 5; Length 353;  
 Best Local Similarity 50.0%; Pred. No. 70;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
 DB 217 IMGICIVTVEV 228

RESULT 12  
 ID Q914D5 PRELIMINARY; PRT; 135 AA.  
 AC Q914D5;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE HYPOTHETICAL PROTEIN PA1203.  
 GN PA1203.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-PA01;  
 RX MEDLINE-20437337; PubMed-10984043;  
 RA Stover C.K., Pham X.-Q.T., Ervin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Labig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004550; AAG04592.1; -;  
 KW Hypothetical protein.  
 SO SEQUENCE 135 AA; 14681 MW; 64DBF1D9549A4818 CRC64;

Query Match 66.1%; Score 37; DB 2; Length 135;  
 Best Local Similarity 54.5%; Pred. No. 45;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 IGICAVTAV 12  
: 1 1 1 1 1 1  
Db 51 LGACTATVAM 61

## RESULT 13

Q9RF47 PRELIMINARY; PRT; 166 AA.  
AC Q9RF47;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
DE LIPOPROTEIN SIGNAL PEPTIDASE (EC 3.4.23.36) (PROLIPOPROTEIN SIGNAL PEPTIDASE) (SIGNAL PEPTIDASE II) (SPASE II).  
GN LSP.  
OS Klebsiella pneumoniae.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Klebsiella.  
OX NCBI\_TaxID=573;  
RN [1]  
RP STRAIN=ATCC13883;  
RC Genova E.G., Richard C.L., Achenbach L.A.;  
RT "Independent Expression and Iron Regulation of the Lsp Gene in Klebsiella pneumoniae."  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY CATALYZES THE REMOVAL OF SIGNAL PEPTIDES FROM PROLIPOPROTEINS (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF N-TERMINAL LEADER SEQUENCES FROM MEMBRANE PROLIPOPROTEINS. HYDROLYSES XAA-XBB-XBB-1-CYS, IN WHICH XAA IS OFTEN GLY OR ALA, AND THE CYS IS ALKYLATED ON SULFUR WITH A DIACYLGLYCERYL GROUP (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC EMBL: AF201388; AAF19640.1; -;  
DR InterPro: IPR001872; -;  
DR Pfam: PF01252; Peptidase\_A8; 1.  
DR PRINTS: PR00781; LIPOSIGPTASE.  
DR PROSITE: PS00855; SPASE II; 1.  
KW Aspartyl protease; Hydrolase; Lipoprotein; Transmembrane.  
SQ SEQUENCE 166 AA; 10410 MW; 474857EDDC085EF CRC64;

Query Match 66.1%; Score 37; DB 2; Length 166;  
Best Local Similarity 77.8%; Pred. No. 54;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 IGICAVTAV 10  
: 1 1 1 1 1 1  
Db 77 IGICVLTAV 85

## RESULT 14

Q45385 PRELIMINARY; PRT; 193 AA.  
AC Q45385;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
DE PUTATIVE GTG START CODON.  
GN BPLK.  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;  
OC Bordetella.  
OX NCBI\_TaxID=520;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BP536;  
RX MEDLINE=96419162; Pubmed=8821935;  
RA Allen A.G., Maskell D.J.;  
RT "The Identification, Cloning and Mutagenesis of a Genetic Locus required for lipopolysaccharide biosynthesis in Bordetella

RT pertussis.";  
RL Mol. Microbiol. 19:37-52(1996).  
DR EMBL: X90711; CA62254.1; -;  
SQ SEQUENCE 193 AA; 21015 MW; C59502AB80618650 CRC64;

Query Match 66.1%; Score 37; DB 2; Length 193;  
Best Local Similarity 75.0%; Pred. No. 61;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 LGICVAV 8  
: 1 1 1 1 1 1  
Db 166 LGVGVAV 173

## RESULT 15

O31603 PRELIMINARY; PRT; 218 AA.  
AC O31603;  
DT 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
DE YJBE PROTEIN.  
GN YJBE.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; Pubmed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borcher S.,  
RA Bottis R., Bourcier L., Brans A., Brann M., Brignell S.C., Bron S.,  
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,  
RA Danisot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,  
RA Ertan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Funa S., Gallier A., Galleron N.,  
RA Gilm S.Y., Glaeser P., Goffeau A., Gohlty E.J., Grandi G.,  
RA Giuseppe G., Guy B.J., Haga K., Haeck J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
RA Medina N., Meliado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terstera P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartot A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzengger T.,  
RA Winters P., Wipet A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the gram-positive bacterium Bacillus subtilis.";  
RL Nature 390:249-256(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Z99110; CAB13008.1; -;  
SQ SEQUENCE 218 AA; 23745 MW; EA254C8B95A90700 CRC64;

Query Match 66.1%; Score 37; DB 2; Length 218;

Tue Nov 13 15:33:07 2001

us-09-412-558-2.rspt

Page 6

Best Local Similarity 58.3%; Pred. No. 68;  
Matches 7; Conservative 3; Mismatches

2; Indels 0; Gaps 0;

QY 1 LIGICVAVTVAI 12  
:|:|:|:|:|  
Db 135 VIGLCVSPVPII 146

Search completed: November 13, 2001, 14:34:01  
Job time: 427 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:26:14 ; Search time 45.32 seconds

(without alignments)  
13.377 Million cell updates/sec

Title: US-09-412-558-1

Sequence: 1 EHMVSYGLRPG 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_0601.\*  
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	100.0	10	2	AA19411
2	63	100.0	10	2	AA19416
3	63	100.0	10	6	AA19422
4	63	100.0	10	7	AA19427
5	63	100.0	10	7	AA19430
6	63	100.0	10	7	AA19436
7	63	100.0	10	8	AA19442
8	63	100.0	10	10	AA19450
9	63	100.0	10	12	AA19453
10	63	100.0	10	13	AA19459
11	63	100.0	10	15	AA19469

12	63	100.0	10	16	AA19497	LHRH peptide, Syn
13	63	100.0	10	16	AA19498	Gonadotropin relea
14	63	100.0	10	16	AA19499	Gonadotropin relea
15	63	100.0	10	17	AA19501	Luteinizing hormon
16	63	100.0	10	17	AA19502	Luteinizing hormon
17	63	100.0	10	18	AA19503	Luteinizing hormon
18	63	100.0	10	18	AA19504	Luteinizing hormon
19	63	100.0	10	19	AA19505	Neutrophil-activat
20	63	100.0	10	20	AA19506	Neutrophil-activat
21	63	100.0	10	20	AA19507	Ubiqutlin fusion p
22	63	100.0	10	20	AA19508	Non-crosslinked pr
23	63	100.0	10	20	AA19509	Amino acid sequenc
24	63	100.0	10	20	AA19510	LHRH peptide fragm
25	63	100.0	10	20	AA19511	Luteinizing hormon
26	63	100.0	10	20	AA19512	Hormone domain of
27	63	100.0	10	20	AA19513	Luteinizing hormon
28	63	100.0	10	21	AA19514	Gonadotropin relea
29	63	100.0	10	21	AA19515	Human LHRH peptide
30	63	100.0	10	21	AA19516	Gonadotropin relea
31	63	100.0	10	21	AA19517	Luteinizing hormon
32	63	100.0	10	21	AA19518	Gonadotropin relea
33	63	100.0	10	21	AA19519	Amino acid sequenc
34	63	100.0	10	21	AA19520	Luteinizing hormon
35	63	100.0	10	21	AA19521	Gonadotropin relea
36	63	100.0	10	21	AA19522	Gonadotropin relea
37	63	100.0	10	21	AA19523	Mammalian releasin
38	63	100.0	10	21	AA19524	Luteinizing hormon
39	63	100.0	10	21	AA19525	LHRH target antige
40	63	100.0	10	21	AA19526	Luteinizing hormon
41	63	100.0	10	21	AA19527	Native mammalian g
42	63	100.0	10	21	AA19528	Luteinizing hormon
43	63	100.0	10	22	AA19529	GHRH monomer pepti
44	63	100.0	10	22	AA19530	Mammalian lutealins
45	63	100.0	10	22	AA19531	Luteinizing Hormon

#### ALIGNMENTS

RESULT 1	
ID	AA19411 standard; peptide; 10 AA.
AC	XX
XX	AA19411:
DT	17-DEC-1992 (first entry)
XX	XX
DE	Luteinizing Hormone Releasing Hormone.
XX	XX
KW	LHRH; Follicle Stimulating Factor; FSH; acne; hirsutism;
KW	dysmenorrhea; precocious puberty; endometriosis; prostate cancer;
KW	benign prostate hypertrophy; mammary tumour.
XX	XX
EH	Key
FT	Modified-site 1 Location/Qualifiers
FT	/label= OTHER
FT	/note= "pyroglutamic acid"
FT	Modified-site 10 /note= "amidated"
XX	XX
PN	BE885308-A.
XX	XX
PD	19-MAR-1981.
XX	XX
PF	23-FEB-1983: 83BE-0468932.
XX	XX
PR	21-SEP-1979: 79FR-0023545.
XX	XX
PA	(ROUS ) ROUSSEL UCLAF.
XX	XX
DR	WPI: 1981-23409D/14 (23409D).
XX	XX
PT	LH-RH, liberating factor for LH and FSH, and its agonists compsn.

PT - used to treat prostate adenocarcinoma, benign hypertrophy of  
the prostate, hirsutism, acne, etc.

Claim 1(a): Page 15; 27pp; French.

PS A composition is claimed containing LHRH or its analogues. The  
CC composition is used to treat prostate adenocarcinoma, benign  
CC hypertrophy of the prostate, endometriosis, dysmenorrhea, hirsutism,  
CC hormone-dependent mammary tumours, for treatment or prevention of  
CC precocious puberty, delaying the onset of puberty and for treating  
CC acne. The compositions may also contain antiandrogens.  
CC See also AAP10412-P10418.  
XX

SQ Sequence 10 AA:

Query Match 100.0%; Score 63; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHMSYGLRPG 10  
| | | | | | | | | |  
DB 1 ehwsyglrpg 10

## RESULT 2

AAP10416  
ID AAP10416 standard; peptide; 10 AA.

XX AAP10416;

DT 17-DEC-1992 (first entry)

XX Luteinising Hormone Releasing Hormone analogue #5.

DE LHRH; Follicle Stimulating Factor; FSH; acne; hirsutism;

XX LHRH; Follicle Stimulating Factor; FSH; acne; hirsutism;

KM dysmenorrhea; precocious puberty; endometriosis; prostate cancer;

KM benign prostate hypertrophy; mammary tumour.

XX Location/Qualifiers

FT Modified-site 1 /label= OTHER

FT Modified-site /note= "pyroglutamic acid"

FT Modified-site /label= OTHER

FT Modified-site /note= "N-alpha-methyl-Leu"

FT Modified-site /note= "amidated or absent, in which case Pro(9)

FT Modified-site /note= "is Pro-NH-C2H5"

FT Modified-site

FT Modified-site

FT Modified-site

FT Modified-site

FT Modified-site

FT Modified-site

FT Modified-site

FT Modified-site

FT Modified-site

FT Modified-site

CC acne. The compositions may also contain antiandrogens.  
CC See AAP10411-P10418.  
XX

SQ Sequence 10 AA:

Query Match 100.0%; Score 63; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHMSYGLRPG 10  
| | | | | | | | | |  
DB 1 ehwsyglrpg 10

## RESULT 3

AAP50222  
ID AAP50222 standard; Protein; 10 AA.

XX AAP50222;

DT 20-JAN-1992 (first entry)

XX Gonadotrophin release stimulating hormone.

XX GnRH; LH-RH; LRF; gonadotrophins; steroids; contraceptive.

XX Synthetic.

XX EPL43573-A.

XX 05-JUN-1985.

XX 05-NOV-1984; 84EP-0307625.

XX 29-NOV-1983; 83US-0556148.

XX 30-AUG-1985; 85US-0771517.

XX (SALK ) SALK INST FOR BIOL STUD.

XX Roeske RW, Rivier JE, Vale WW;

XX WPI; 1985-136434/23.

XX New GnRH antagonist peptide(s) - useful as inhibitors of

XX gonadotropin(s) and/or steroid(s) for contraceptive use.

XX Disclosure: Page 1; 20pp; English.

XX The claimed peptide antagonists inhibit the release of gonadotrophins

XX and/or steroids. They are antagonistic to GnRH, inhibit ovulation, and

XX may cause resorption of a fertilised egg if administered shortly after

XX absorption. The peptides also have utility in male contraception, and

XX in treatment of precocious puberty, hormone dependent neoplasia,

XX dysmenorrhoea and endometriosis.

XX Sequence 10 AA:

OY 1 EHMSYGLRPG 10  
| | | | | | | | | |  
DB 1 ehwsyglrpg 10

## RESULT 4

AAP60127  
ID AAP60127 standard; Peptide; 10 AA.

XX AAP60127;



XX 12-JUN-1991 (first entry)  
 XX Gonadoliberin antagonist.  
 DE Gonadoliberin antagonist.  
 KW Gonadoliberin antagonist; contraceptive; antitumor.  
 XX EP201260-A.  
 XX 12-NOV-1986.  
 XX 28-APR-1986; 86EP-0303210.  
 XX 09-MAY-1985; 85US-0732531.  
 XX (SALK ) SALK INST FOR BIOL STUD.  
 PA Rivier JEF, Varga JT, Hagler AT, Struthers RS, Perrin MH;  
 PI Rivier CL, Vale MW;  
 XX WPI; 1986-299774/46.  
 DR New peptide gonadotropin releasing hormone antagonists - useful  
 PT esp. as contraceptives, for treating early puberty,  
 PT hormone-dependent neoplasms etc.  
 PS Disclosure: Page 1; 33pp: English.  
 XX The decapeptide encodes a gonadoliberin antagonist, which may be  
 CC used as a male contraceptive and as an antitumour (against steroid-  
 CC dependent tumours).  
 XX Sequence 10 AA;  
 SQ

Query Match 100.0%; Score 63; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 DB 1 ehwsyglrpg 10

## RESULT 5

AAP61403  
 ID AAP61403 standard; protein; 10 AA.  
 AC AAP61403;  
 XX

DT 04-AUG-1991 (first entry)  
 XX Gonadotropin releasing hormone.  
 DE Gonadotropin releasing hormone.  
 XX Gonadotropin releasing hormone; analogue; peptide synthesis;  
 KW ovulation; veterinary medicine; fertility;  
 XX DD232500-A.  
 XX 29-JAN-1986.  
 PD 08-MAY-1984; 84DD-0262804.  
 PF 08-MAY-1984; 84DD-0262804.  
 PR 08-MAY-1984; 84DD-0262804.  
 XX (DEAK ) AKAD WISSENSCHAFT DDR.  
 PA Kaufmann KD, Dolling R, Handel L;  
 PI WPI; 1986-137868/22.  
 DR Prepn. of gonadotropin liberating hormone and analogues - by  
 PT multistage rapid peptide synthesis in soln. without isolating

PT Intermediates  
 XX Disclosure: page 7; 8pp; german.  
 PS The gonadotropin releasing hormone and its analogues are prep. by a  
 CC new multistage rapid peptide synthesis method in soln., where the  
 CC intermediates are not isolated. The process is rapid and gives very  
 CC pure peptide quickly and using little equipment. The peptide can be  
 CC used in veterinary medicine to synchronise ovulation in large animal  
 CC herds, and in human medicine in the treatment of fertility disorders.  
 XX Sequence 10 AA;  
 SQ

Query Match 100.0%; Score 63; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 DB 1 ehwsyglrpg 10

## RESULT 6

AAP60576  
 ID AAP60576 standard; protein; 10 AA.  
 AC AAP60576;  
 XX

DT 27-OCT-1991 (first entry)  
 XX Novel decapeptide with LHRH inhibition activity.  
 DE Lutenising hormone releasing hormone activity.  
 XX Synthetic.  
 OS JP61210098-A.  
 PN 18-SEP-1986.  
 PD 23-AUG-1985; 85JP-0185616.  
 PF 23-AUG-1984; 84US-0643643.  
 PR (TULA-) ADMIN TULANE EDUCAT.  
 PA (TULA ) TULANE E FUND ADMINISTR.  
 XX WPI; 1986-321434/49.  
 DR Deca:peptide - inhibits LH-RH hormone release activity.  
 PT Disclosure: Page 990; 5pp; Japanese.  
 PS Peptide inhibits the release of lutenising hormone releasing hormone.  
 CC See also AAP60575.  
 XX Sequence 10 AA;  
 SQ

Query Match 100.0%; Score 63; DB 7; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 DB 1 ehwsyglrpg 10

## RESULT 7

AAP70922  
 ID AAP70922 standard; peptide; 10 AA.  
 XX



XX	(INRM ) INSEPM INST NAT SANTE.
PA	
XX	
PI	Gautron J, Patou E, Kordon C, Bauer K;
XX	
DR	WPI; 1991-339753/46.
XX	
PT	New peptide homologous with luteinising hormone-releasing hormone
PT	- used to treat gynaecological conditions, cancer of gonads and
PT	sec. sexual organs, psychiatric conditions and in assays
XX	
PS	Claim 3; Page 50; 83pp; French.
XX	
CC	The C-terminal residue (Gly-Co-NH <sub>2</sub> ) can be replaced by ethylamide,
CC	This peptide and fragments of it (i.e. amino acids 4-10, 5-10, 6-10
CC	and 7-10) are agonists and antagonists of LHRH. They are useful for
CC	treating e.g. precocious or delayed puberty, psychiatric disorders
CC	esp. those of the libido or sexual aggression, etc. In addition they
CC	are useful for functional exploration of the hypothalamus-hypophyseal
CC	axis and for radiomunological or biological assay (of LH, FSH and
CC	steroid levels) in biological fluids and biopsy samples.
XX	
SQ	Sequence 10 AA:
Query Match	100.0%; Score 63; DB 12; Length 10;
Best Local Similarity	100.0%; Pred. No. 0.00012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 EHWSYGLRPG 10 
DB	1 ehwsyglrpg 10
RESULT 10	
AAR26819	
ID	AAR26819 standard; peptide; 10 AA.
XX	
AC	AAR26819;
XX	
DT	10-FEB-1993 (first entry)
XX	
DE	LH releasing hormone antagonists.
XX	
KW	Luteinising hormone; LHRH; hypothalamic; antiovaratory; tumours;
KW	antineoplastic; precocious puberty; ovulation; contraceptive.
XX	
OS	Synthetic.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 1 /label= pcju
FT	Modified-site 10 /note= "amidated"
PN	W09213883-A.
XX	
PD	20-AUG-1992.
XX	
PJ	29-JAN-1992; 92MO-US00776.
XX	
PR	30-JAN-1991; 91US-0647786.
XX	
PA	(TULA ) TULANE EDUCATIONAL FUND.
XX	
PI	Janaky T, Juhasz A, Schally AV;
XX	
DR	WPI; 1992-299984/36.
XX	
PT	New decapeptide luteinising hormone-releasing hormone
PT	antagonists - for treating precocious puberty, hormone dependent
PT	tumours, endometritis, cystic diseases; also as contraceptive

PS Disclosure: Page 1, 43pp; English.

XX The decapeptides is an antagonistic analogue of hypothalamic LHRH  
CC which possesses high anti-ovulatory and antineoplastic activity, is  
CC free of anaphylactoid side effects and is believed to be free of  
CC endemotogenic effects. The peptide may be used to treat precocious  
CC puberty, hormone dependent tumours, e.g. malignant and benign  
CC prostate tumours, e.g. secondary amenorrhoea, endometriosis and  
CC ovarian and mammary cystic diseases. The peptide is also useful  
CC for regulating ovulation e.g. as pre-coital or post-coital  
CC contraceptives, for synchronising oestrus in livestock and for  
CC improving the "rhythm" method. It is also useful for regulating  
CC the human menopausal gonadotropin, follicle stimulating and LH levels  
CC during premenopausal and postmenopausal periods. As it suppresses  
CC the spermatogenesis and testosterone levels in males, it may be of  
CC potential use for male contraception.  
CC See also AAR26818, AAR29046-7.

XX  
XX Sequence 10 AA:

SO

Query Match 100.0%; Score 63; DB 13; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.00012;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHWSTYGLRPG 10  
          | | | | | | | | | |  
DB 1 EHWSTYGLRPG 10

RESULT 11  
AAR62689  
ID AAR62689 standard; peptide: 10 AA.

XX  
XX AAR62689;

AC  
DT 10-SEP-1995 (first entry)

XX  
XX LHRH hapten for attachment to universal immune stimulator.

DE  
XX  
XX Helper T cell epitope; universal immune stimulator; invasive; hapten;  
KW vaccine; LHRH; luteinising hormone releasing hormone; prostate;  
KW androgen-dependent carcinoma; antitumour; infertility.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO9425060-A.  
PN  
XX 10-NOV-1994.  
PD  
XX  
XX 28-APR-1994; 94WO-US04832.  
PF  
XX  
XX 27-APR-1993; 93US-0057166.  
PR  
XX 14-APR-1994; 94US-0229275.  
PP  
XX  
XX (LADD/) LADD A E.  
PA (WANG/) WANG C Y.  
PA (ZAMB/) ZAMB T.  
XX  
XX Ladd AE, Wang CY, Zamb T;  
PI  
XX  
XX WPI: 1994-357910/44.  
DR  
XX  
XX Immunogenic luteinising hormone releasing hormone peptide(s) -  
PT that suppress LHRH activity in males and females  
XX  
XX Claim 6; Page 104; 213pp; English.

PS  
XX Synthetic immunogenic peptides are provided in which a universal immune  
CC stimulator is linked to a peptide or protein hapten containing B cell  
CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes  
CC potent immune responses to the coupled peptide or protein. The  
CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)



CC analogues may also exert effects on the autonomic nervous system. The  
 CC GnRH analogues are administered by injection (which may be intravenous,  
 CC subcutaneous or intramuscular), or by a drug delivery system. The drug  
 CC delivery system can comprise a drug implant with timed release, a nasal  
 CC spray or an injection of a long-lasting depo form. This method is used  
 CC to alleviate symptoms such as nausea, vomiting, abdominal pain and  
 CC altered bowel habits. The sequences can be used to treat motility  
 CC disorders in a wide variety of other diseases including irritable bowel  
 CC syndrome, diabetes, scleroderma and Parkinson's disease.

XX Sequence 10 AA:

Query Match 100.0%; Score 63; DB 16; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 |||||  
 Db 1 ehwsyglrpg 10

RESULT 14

AAK75152  
 ID AAK75152 standard; peptide: 10 AA.

XX AAK75152;

DT 19-DEC-1995 (first entry)

DE Gonadotropin releasing hormone.

KM Gonadotropin releasing hormone; GnRH; gonadolibertin; reproduction;

transgenic animal; transgenic fish; transgenic fowl.

XX Mammalia.

OS WO9512309-A1.

XX PD 11-MAY-1995.

XX PF 04-NOV-1994; 94WO-US12763.

XX PR 05-NOV-1993; 93US-0147771.

XX PA (STRP ) UNIV LELAND STANFORD JUNIOR.

XX PA (UYOR-) UNIV OREGON HEALTH SCI.

XX PI (UYOR-) UNIV OREGON STATE.

XX PI Adelman JP, Fernald RD;

XX DR WPI: 1995-185526/24.

XX PT New gonadotropin releasing hormone preprohormone DNA - used to  
 PT develop prods. for regulation of reproductive function and diagnosis  
 PT of reproductive capacity and disease

XX PS Disclosure: Fig.1a; 85pp; English.

CC 8 Different forms of GnRH (given in AAK75152-59) have previously  
 CC been isolated from vertebrate species. A precursor for an  
 CC additional form of GnRH, (Ser8)-GnRH (AAK75151), has now been  
 CC obtd.

XX SQ Sequence 10 AA:

Query Match 100.0%; Score 63; DB 16; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 |||||

Db 1 ehwsyglrpg 10

RESULT 15

AAW65201  
 ID AAW65201 standard; peptide: 10 AA.

XX AAW65201;

XX DT 02-OCT-1998 (first entry)

XX DE Luteinising hormone-releasing hormone (LH-RH).

XX KM Bradykinin; N-benzylglycine; agonist; receptor study; antagonist;

XX KM achiral; analgesic; luteinising hormone-releasing hormone; LHRH;

XX KW gonadolibertin.

XX OS Synthetic.

XX FT Key

FT Modified-site 1 Location/Qualifiers  
 FT Modified-site 10 /note="Pyroglutamic acid"  
 FT Modified-site 10 /note="C-terminal amide"

XX PN US5527882-A.

XX PD 18-JUN-1996.

XX PF 07-NOV-1994; 94US-0335202.

XX PR 07-JUL-1989; 89US-0376839.

XX PR 16-SEP-1992; 92US-0945664.

XX PR 07-NOV-1994; 94US-0335202.

XX PA (REGC ) UNIV CALIFORNIA.

XX PI Mitchell AR, Young JD;

XX DR WPI: 1996-299898/30.

XX PT New bradykinin analogues contg. N-benzyl-glycine - useful as  
 PT bradykinin agonists or antagonists, useful e.g. as analgesics

XX PS Disclosure: Columns 11-12; 15pp; English.

CC The invention relates to the obtaining of a potent agonist or antagonist  
 CC peptide by the replacement of selected amino acids with synthetic  
 CC achiral amino acids. The present sequence represents a luteinising  
 CC hormone-releasing hormone (LHRH).

XX SQ Sequence 10 AA:

Query Match 100.0%; Score 63; DB 17; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 |||||  
 Db 1 ehwsyglrpg 10

Search completed: November 13, 2001, 14:27:41  
 Job time: 87 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:26:14 : Search time 26.87 Seconds  
(without alignments)  
8.375 Million cell updates/sec

Title: US-09-412-558-1  
Perfect score: 63  
Sequence: 1 EHMSYGLRPG 10

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents-AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/CTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	100.0	10 1	US-07-714-540-9	Sequence 9, Appl
2	63	100.0	10 1	US-07-690-983D-2	Sequence 2, Appl
3	63	100.0	10 1	US-07-690-983D-32	Sequence 32, Appl
4	63	100.0	10 1	US-08-343-883-1	Sequence 5, Appl
5	63	100.0	10 1	US-08-000-931-5	Sequence 22, Appl
6	63	100.0	10 1	US-08-428-488-22	Sequence 11, Appl
7	63	100.0	10 1	US-08-341-219-11	Sequence 22, Appl
8	63	100.0	10 1	US-08-453-588-22	Sequence 11, Appl
9	63	100.0	10 1	US-08-591-917-1	Sequence 22, Appl
10	63	100.0	10 1	US-08-446-692-1	Sequence 1, Appl
11	63	100.0	10 2	US-08-796-598-6	Sequence 1, Appl
12	63	100.0	10 2	US-08-694-865-18	Sequence 18, Appl
13	63	100.0	10 2	US-08-488-351A-1	Sequence 1, Appl
14	63	100.0	10 2	US-08-480-494B-1	Sequence 1, Appl
15	63	100.0	10 2	US-08-447-175A-6	Sequence 6, Appl
16	63	100.0	10 3	US-08-521-079-22	Sequence 22, Appl
17	63	100.0	10 3	US-09-124-491-18	Sequence 18, Appl
18	63	100.0	10 3	US-09-100-414B-77	Sequence 77, Appl
19	63	100.0	10 3	US-08-927-128-13	Sequence 13, Appl
20	63	100.0	10 4	US-08-912-314A-11	Sequence 11, Appl
21	63	100.0	10 4	US-09-303-323-77	Sequence 77, Appl
22	63	100.0	10 6	5168061-1	Patent No. 5168061
23	63	100.0	10 6	5169865-10	Patent No. 5169865
24	63	100.0	10 6	5169935-1	Patent No. 5169935
25	63	100.0	10 6	5488036-1	Patent No. 5488036
26	63	100.0	10 6	5492893-1	Patent No. 5492893
27	63	100.0	10 6	5492893-1	Patent No. 5492893

28	63	100.0	12 1	US-08-302-915-2	Sequence 2, Appl
29	63	100.0	14 1	US-07-690-983D-22	Sequence 22, Appl
30	63	100.0	14 1	US-07-690-983D-24	Sequence 24, Appl
31	63	100.0	14 1	US-07-690-983D-26	Sequence 26, Appl
32	63	100.0	14 1	US-07-690-983D-30	Sequence 30, Appl
33	63	100.0	16 1	US-07-690-983D-14	Sequence 14, Appl
34	63	100.0	17 1	US-07-690-983D-16	Sequence 16, Appl
35	63	100.0	17 1	US-07-690-983D-18	Sequence 18, Appl
36	63	100.0	18 1	US-07-690-983D-20	Sequence 20, Appl
37	63	100.0	18 1	US-07-690-983D-28	Sequence 28, Appl
38	63	100.0	20 1	US-07-690-983D-40	Sequence 40, Appl
39	63	100.0	24 1	US-07-690-983D-43	Sequence 43, Appl
40	63	100.0	25 1	US-08-446-692-12	Sequence 12, Appl
41	63	100.0	25 1	US-08-446-692-17	Sequence 17, Appl
42	63	100.0	25 2	US-08-488-351A-12	Sequence 12, Appl
43	63	100.0	25 2	US-08-488-351A-17	Sequence 17, Appl
44	63	100.0	26 1	US-08-446-692-29	Sequence 29, Appl
45	63	100.0	26 2	US-08-488-351A-29	Sequence 29, Appl

## ALIGNMENTS

RESULT 1  
US-07-714-540-9  
: Sequence 9, Application US/07714540  
: Patent No. 5262521  
: GENERAL INFORMATION:  
: APPLICANT: Almqvist, Ronald G.  
: APPLICANT: Toll, Lawrence  
: TITLE OF INVENTION: ISOLATED ATRIAL PEPTIDE-DEGRADING  
: ENZYME AND NOVEL COMPOUNDS USEFUL AS INHIBITORS THEREOF  
: NUMBER OF SEQUENCES: 13  
: CORRESPONDENCE ADDRESSES:  
: ADDRESSEE: Irell & Manella  
: STREET: 545 Middlefield Road, Suite 200  
: CITY: Menlo Park  
: STATE: California  
: COUNTRY: USA  
: ZIP: 94025  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/07714,540  
: FILING DATE: 19910607  
: CLASSIFICATION: 530  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Reed, Dianne E.  
: REGISTRATION NUMBER: 31,292  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 415-327-7250  
: TELEFAX: 415-327-2951  
: TELEX: 706141  
: INFORMATION FOR SEQ. ID NO: 9:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 10 amino acids  
: TYPE: AMINO ACID  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: protein  
: US-07-714-540-9

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0.  
QY 1 EHMSYGLRPG 10  
- |||||

Db 1 EHWSYGLRPG 10

RESULT 2  
US-07-690-983D-2  
; Sequence 2, Application US/07690983D  
; Patent No. 5403586  
; GENERAL INFORMATION:  
; APPLICANT: RUSSELL-JONES, Gregory J.  
; APPLICANT: STEWART, Andrew G.  
; APPLICANT: TSONIS, Con G.  
; TITLE OF INVENTION: FUSION PROTEINS  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/690,983D  
; FILING DATE: 25-JUN-1991  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU90/00373  
; FILING DATE: 24-AUG-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 16786/148 CHAC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-07-690-983D-2

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
Db 1 EHWSYGLRPG 10

RESULT 3  
US-07-690-983D-32  
; Sequence 32, Application US/07690983D  
; Patent No. 5403586  
; GENERAL INFORMATION:  
; APPLICANT: RUSSELL-JONES, Gregory J.  
; APPLICANT: STEWART, Andrew G.  
; APPLICANT: TSONIS, Con G.  
; TITLE OF INVENTION: FUSION PROTEINS  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W.  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/690,983D  
; FILING DATE: 25-JUN-1991  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU90/00373  
; FILING DATE: 24-AUG-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 16786/148 CHAC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-690-983D-32

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
Db 1 EHWSYGLRPG 10

RESULT 4  
US-08-343-883-1  
; Sequence 1, Application US/08343883  
; Patent No. 5573767  
; GENERAL INFORMATION:  
; APPLICANT: Dufour, Raymond J.  
; APPLICANT: Roulet, Claude J.M.  
; APPLICANT: Chouvet, Claire D.  
; APPLICANT: Bonneau, Michel B.  
; TITLE OF INVENTION: Method for improving the organoleptic  
; TITLE OF INVENTION: qualities of the meat from uncastated male domestic  
; TITLE OF INVENTION: animals, vaccines which are usable in this method, new  
; TITLE OF INVENTION: peptide, in particular for producing these vaccines...  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Larson and Taylor  
; STREET: 727 Twenty-Third Street, South  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/343,883  
; FILING DATE: 17-NOV-1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/946,495  
; FILING DATE: 09-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 9102513  
; FILING DATE: 01-MAR-1991  
; PRIOR APPLICATION DATA:



APPLICATION NUMBER: FR 9115289  
FILING DATE: 10-DEC-1991  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 10  
OTHER INFORMATION: /label= NH2  
OTHER INFORMATION: /note= "amidated glycine"  
FEATURE:  
NAME/KEY: Peptide  
LOCATION: 1  
OTHER INFORMATION: /label= pyro  
OTHER INFORMATION: /note= "pyroglutamic acid"  
PUBLICATION INFORMATION:  
AUTHORS: Matsuo, H.  
AUTHORS: Baba, Y.  
AUTHORS: G. Nair, R. M.  
AUTHORS: Arimura, A. V.  
AUTHORS: Schally, A. V.  
TITLE: Structure of the porcine LH- and  
TITLE: FSH-releasing hormone. I. The proposed amino acid  
TITLE: sequence.  
JOURNAL: Biochem. Biophys. Res. Commun.  
VOLUME: 43  
ISSUE: 6  
PAGES: 1334-1339  
DATE: 1971  
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 10  
US-08-343-883-1

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
DB 1 EHWSYGLRPG 10

RESULT 5  
US-08-000-931-5  
Sequence 5; Application US/08000931  
Patent No. 5578477  
GENERAL INFORMATION:  
APPLICANT: Tamanoi Dr., Fuyuhiko  
TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF  
NUMBER OF SEQUENCES: 10  
TITLE OF INVENTION: INHIBITORS OF PROTEIN FARNESYLTRANSFERASE  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/000,931  
FILING DATE: 05-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 64098/102/ARDE

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-000-931-5

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
DB 1 EHWSYGLRPG 10

RESULT 6  
US-08-428-488-22  
Sequence 22; Application US/08428488  
Patent No. 5624894  
GENERAL INFORMATION:  
APPLICANT: BODOR, Nicholas S.  
TITLE OF INVENTION: BRAIN-ENHANCED DELIVERY OF NEUROACTIVE  
TITLE OF INVENTION: PEPTIDES BY SEQUENTIAL METABOLISM  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/428,488  
FILING DATE: 27-APR-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Baumeister, Mary Katherine  
REGISTRATION NUMBER: 26,254  
REFERENCE/DOCKET NUMBER: 028724-087  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /note= "Position 1 = p-Glu."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 10  
OTHER INFORMATION: /note= "Position 10 = Gly-NH2."  
US-08-428-488-22

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
Db 1 EHWSYGLRPG 10

RESULT 7  
US-08-341-219-11  
; Sequence 11, Application US/08341219  
; Patent No. 5643877  
; GENERAL INFORMATION:  
; APPLICANT: Zohar, Y.  
; APPLICANT: Rivier, J.  
; APPLICANT: Powell, J.  
; APPLICANT: Sherwood, N.  
; APPLICANT: Gotthelf, Y.  
; TITLE OF INVENTION: Compounds and Methods for Controlling  
; TITLE OF INVENTION: Reproduction in Fish  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/341,219  
; FILING DATE: 05-DEC-1994  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30742  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 1  
; OTHER INFORMATION: /product= "OTHER"  
; OTHER INFORMATION: /label= Glu1  
; OTHER INFORMATION: /note= "pyroglutamic acid"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 10  
; OTHER INFORMATION: /product= "OTHER"  
; OTHER INFORMATION: /label= Gly10  
; OTHER INFORMATION: /note= "amidated"  
US-08-341-219-11

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
Db 1 EHWSYGLRPG 10

RESULT 8  
US-08-453-588-22  
; Sequence 22, Application US/08453588  
; Patent No. 5684145  
; GENERAL INFORMATION:  
; APPLICANT: Anna van der Zee, Irma Marianne van Die,  
; APPLICANT: Willem Pieter Martin Hoekstra,  
; APPLICANT: Josephus Theodorus Gijlen.  
; TITLE OF INVENTION: Carrier system against GNRH  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Akzo No. 5684145el Patent Department  
; STREET: 1300 Piccard Drive, Suite 206  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: U.S.A.  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/453,588  
; FILING DATE: 30-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/078,661  
; FILING DATE: 16-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mary E. Gormley  
; REGISTRATION NUMBER: 34,409  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 258-5200  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Glu at position 1 is pyroglutamic acid  
US-08-453-588-22

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
Db 1 EHWSYGLRPG 10

RESULT 9  
US-08-591-917-1  
; Sequence 1, Application US/08591917  
; Patent No. 570764  
; GENERAL INFORMATION:  
; APPLICANT: Nett, Torrance M  
; APPLICANT: Glode, Leonard Michael  
; TITLE OF INVENTION: A METHOD FOR TREATING CANCER  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross & McIntosh  
; STREET: 1700 Lincoln Street, Suite 3500  
; CITY: Denver  
; STATE: Colorado

COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/591,917  
FILING DATE: 26-JAN-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Kovarik, Joseph E.  
REGISTRATION NUMBER: 33,005  
REFERENCE/DOCKET NUMBER: 2730-3-2-1-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-591-917-1

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHMSYGLRPG 10  
DB 1 EHMSYGLRPG 10

RESULT 10  
US-08-446-692-1  
Sequence 1, Application US/08446692  
Patent No. 5759551  
GENERAL INFORMATION:  
APPLICANT: Ladd, Anna  
APPLICANT: Wang, Chang Y1  
APPLICANT: zamb, Timothy  
TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C.H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10154-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446,692  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria C.H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-446-692-1

Query Match 100.0%; Score 63; DB 1; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHMSYGLRPG 10  
DB 1 EHMSYGLRPG 10

RESULT 11  
US-08-796-598-6  
Sequence 6, Application US/08796598  
Patent No. 5827659  
GENERAL INFORMATION:  
APPLICANT: PATTERSON, DALE H.  
APPLICANT: TARR, GEORGE E.  
TITLE OF INVENTION: METHODS AND APPARATUS FOR SEQUENCING  
TITLE OF INVENTION: POLYMERS USING MASS SPECTROMETRY.  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patent Administrator - Testa, Hurwitz &  
ADDRESSEE: Thibault  
STREET: High Street Tower, 125 High Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/796,598  
FILING DATE: 07-FEB-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/446,055  
FILING DATE: 19-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: FLYNN Esq., Kerry A.  
REGISTRATION NUMBER: 33,693  
REFERENCE/DOCKET NUMBER: SYP-115  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-796-598-6

Query Match 100.0%; Score 63; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHMSYGLRPG 10  
DB 1 EHMSYGLRPG 10

RESULT 12  
US-08-694-865-18

Sequence 18, Application us/08694865  
Patent No. 5837268  
GENERAL INFORMATION:  
APPLICANT: POTTER, ANDREW A.  
APPLICANT: MANNIS, JOHN G.  
TITLE OF INVENTION: GNRH-LEUKOTOXIN CHIMERAS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: REED & ROBINS LLP  
STREET: 285 HAMILTON AVENUE, SUITE 200  
CITY: PALO ALTO  
STATE: CA  
COUNTRY: USA  
ZIP: 94301  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/694,865  
FILING DATE: 09-AUG-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: MCCracken, THOMAS P.  
REGISTRATION NUMBER: 38,548  
REFERENCE/DOCKET NUMBER: 9001-0016.22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)327-3400  
TELEFAX: (415)327-3231  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /note="This position is pyroglu."  
US-08-694-865-18

Query Match 100.0%; Score 63; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. NO. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSTGLRPG 10  
DB 1 EHWSTGLRPG 10

RESULT 13  
US-08-488-351A-1  
Sequence 1, Application US/08488351A  
Patent No. 5843446  
GENERAL INFORMATION:  
APPLICANT: Ladd, Anna  
APPLICANT: Wang, Chang YI  
APPLICANT: Zam, Timothy  
TITLE OF INVENTION: Immunogenic LHRH peptide constructs  
TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines  
NUMBER OF SEQUENCES: 114  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Maria C. H. Lin  
STREET: 345 Park Avenue  
CITY: New York  
STATE: NY  
COUNTRY: US  
ZIP: 10154-0053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,351A  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/446,692  
FILING DATE: 7-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/229,275  
FILING DATE: 14-APR-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/057,166  
FILING DATE: 27-APR-1992  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Maria C. H. Lin  
REGISTRATION NUMBER: 29,323  
REFERENCE/DOCKET NUMBER: 1151-4146 US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)415-8745  
TELEFAX: (516)751-6849  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-488-351A-1

Query Match 100.0%; Score 63; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. NO. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSTGLRPG 10  
DB 1 EHWSTGLRPG 10

RESULT 14  
US-08-480-494B-1  
Sequence 1, Application US/08480494B  
Patent No. 5843901  
GENERAL INFORMATION:  
APPLICANT: Roeske, Roger W.  
TITLE OF INVENTION: LHRH Antagonist Peptides  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,494B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: DeConti, Giulio A.  
REGISTRATION NUMBER: 31,503  
REFERENCE/DOCKET NUMBER: PPI-007  
TELECOMMUNICATION INFORMATION:

Search completed: November 13, 2001, 14:26:49  
Job time: 35 sec

TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-480-494B-1

Query Match 100.0%; Score 63; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
|||||

Db 1 EHWSYGLRPG 10

## RESULT 15

US-08-447-175A-6  
Sequence 6, Application US/08447175A  
Patent No. 5869240  
GENERAL INFORMATION:  
APPLICANT: PATTERSON, DALE H.  
TITLE OF INVENTION: METHODS AND APPARATUS FOR SPOURING  
TITLE OF INVENTION: POLYMERS WITH A STATISTICAL CERTAINTY USING MASS  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patent Administrator - Testa, Hurwitz &  
ADDRESSEE: Thibault, LLP  
STREET: High Street Tower, 125 High Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,175A  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 422  
ATTORNEY/AGENT INFORMATION:  
NAME: RAUSCHENBACH, KURT  
REGISTRATION NUMBER: 40,137  
REFERENCE/DOCKET NUMBER: SYP-114  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 248-7000  
TELEFAX: (617) 248-7100  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-447-175A-6

Query Match 100.0%; Score 63; DB 2; Length 10;  
Best Local Similarity 100.0%; Pred. No. 7e-05;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
|||||

Db 1 EHWSYGLRPG 10

Tue Nov 13 15:32:57 2001

us-09-412-558-1.ra1

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:26:14 ; Search time 30.29 Seconds  
(without alignments)  
25.148 Million cell updates/sec

Title: US-09-412-558-1

Perfect score: 63

Sequence: 1 EHMSYGLRPG 10

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	95.2	10	1 RHSPG	gonadolibetin - pi
2	60	95.2	10	1 RHSPG	gonadolibetin - sh
3	60	95.2	67	2 I78541	gonadolibetin prec
4	60	95.2	89	2 I51423	gonadolibetin prec
5	60	95.2	90	1 RHMSG	gonadolibetin prec
6	60	95.2	92	1 RHMSG	gonadolibetin prec
7	60	95.2	92	1 RHMSG	gonadolibetin prec
8	56	88.9	10	1 RHAG1	gonadolibetin I -
9	56	88.9	92	2 I50544	gonadolibetin I pr
10	54	85.7	98	2 I50739	gonadolibetin-relea
11	50	79.4	80	1 RHID15	gonadolibetin I pr
12	50	79.4	91	2 JC7393	medaka-type gonado
13	47	74.6	10	2 A21114	gonadolibetin - ch
14	47	74.6	74	2 I51092	gonadolibetin-relea
15	47	74.6	82	2 I51180	gonadolibetin-relea
16	47	74.6	82	2 I51355	gonadolibetin-relea
17	47	74.6	82	2 I51365	gonadolibetin-relea
18	47	74.6	82	2 I51331	gonadolibetin-relea
19	47	74.6	90	2 JC7395	salmon-type gonado
20	47	74.6	90	2 A23735	gonadolibetin prec
21	47	74.6	90	2 I51095	gonadolibetin prec
22	42	66.7	10	1 RHAG2	gonadolibetin II -
23	42	66.7	10	1 A61126	gonadolibetin II -
24	42	66.7	10	2 A49187	gonadolibetin-relea
25	42	66.7	10	2 A46030	gonadolibetin I -
26	42	66.7	10	2 B46030	gonadolibetin II -
27	42	66.7	85	2 JC7394	chicken-II-type go
28	42	66.7	85	2 A53453	gonadolibetin II p
29	42	66.7	86	1 RHID25	gonadolibetin II p

30	42	66.7	828	2 T08556	hypothetical prote
31	41	65.1	316	2 A53440	aldose reductase h
32	41	65.1	532	2 T32849	hypothetical prote
33	40	63.5	551	2 E64728	yabn protein - Esc
34	40	63.5	552	2 B85489	probable transport
35	40	63.5	565	2 B82443	conserved hypothet
36	40	63.5	584	2 J01229	cellulase (EC 3.2.
37	40	63.5	1000	2 C82630	serine proteinase
38	39	61.9	345	2 A58519	hypothetical 345 p
39	39	61.9	488	2 C72710	probable tmu prote
40	39	61.9	417	2 T33827	hypothetical prote
41	39	61.9	501	2 T32848	hypothetical prote
42	39	61.9	508	2 T01937	hypothetical prote
43	39	61.9	1444	2 T18856	angiogenesis inhnb
44	38	60.3	161	2 D84472	hypothetical prote
45	38	60.3	293	2 G72699	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

RHSPG gonadolibetin - pig

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 13-Jul-1981 #sequence\_revision 13-Jul-1981 #text\_change 18-Mar-1997

C:Accession: A01411

R:Baba, Y.; Matsuo, H.; Schally, A.V.

Biochem. Biophys. Res. Commun. 44, 459-463, 1971

A:Title: Structure of the porcine LH- and FSH-releasing hormone. II. Confirmation of

A:Reference number: A90172; MWID:72114303

A:Accession: A01411

A:Molecule type: protein

A:Residues: 1-10 <BAB>

R:Matsuo, H.; Arimura, A.; Nair, R.M.G.; Schally, A.V.

Biochem. Biophys. Res. Commun. 45, 822-827, 1971

A:Title: Synthesis of the porcine LH- and FSH-releasing hormone by the solid-phase me

A:Reference number: A90176; MWID:72065376

A:Contents: annotation: synthesis

A>Note: the synthetic and natural hormones have the same physicochemical and biologic

R:Baba, Y.; Arimura, A.; Schally, A.V.

Biochem. Biophys. Res. Commun. 45, 483-487, 1971

A:Title: On the tryptophan residue in porcine LH and FSH-releasing hormone.

A:Reference number: A90175; MWID:72117544

A:Contents: annotation

A>Note: Trp-3 appears to be essential for biological activity

C:Comment: This hypothalamic hormone stimulates the secretion of both luteinizing and

C:Superfamily: gonadolibetin

C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid

F:1/Modified site: pyroglutamate carboxylic acid (Gln) #status experimental

F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 95.2% Score 60; DB 1; Length 10;

Best Local Similarity 90.0%; Pred. No. 0.00022;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EHMSYGLRPG 10  
Db 1 QHMSYGLRPG 10

##### RESULT 2

RHSPG gonadolibetin - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 18-Mar-1997

C:Accession: A93780; A01411

R:Burgus, R.; Butcher, M.; Amoss, M.; Ling, N.; Monahan, M.; Rivier, J.; Fellows, R.;

Proc. Natl. Acad. Sci. U.S.A. 69, 278-282, 1972

A:Title: Primary structure of the ovine hypothalamic luteinizing hormone-releasing fa

A:Reference number: A93780; MWID:72094314

A:Molecule type: protein  
 A:Residues: 1-10 <R0R>  
 A>Note: the natural and synthetic hormones have the same biological activity  
 C:Comment: This hypothalamic hormone stimulates the secretion of both luteinizing and fo  
 C:Superfamily: gonadoliberin  
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglyutamic acid  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 95.2%; Score 60; DB 1; Length 10;  
 Best Local Similarity 90.0%; Pred. No. 0.00022;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EHWSYGLRPG 10  
 Db 1 QHWSYGLRPG 10

RESULT 3  
 178541  
 gonadoliberin precursor - rhesus macaque (fragment)  
 N:Alternate names: luteinizing hormone releasing hormone  
 C:Species: Macaca mulatta (rhesus macaque)  
 C>Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 16-Jul-1999  
 C:Accession: 178541  
 R:Ma, Y.J.; Costa, M.E.; Ojeda, S.R.  
 Neuroendocrinology 60, 346-359, 1994  
 A>Title: Developmental expression of the genes encoding transforming growth factor alpha  
 A:Reference number: 158134; MUID:95124501  
 A:Accession: 178541  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-67 <RES>  
 A:Cross-references: GB:S75918; NID:g912831; PIDN:AAB33096.1; PID:g912832  
 C:Superfamily: gonadoliberin

Query Match 95.2%; Score 60; DB 2; Length 67;  
 Best Local Similarity 90.0%; Pred. No. 0.0017;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
 Db 6 QHWSYGLRPG 15  
 RESULT 4  
 151423  
 gonadoliberin precursor - African clawed frog  
 N:Alternate names: luteinizing hormone releasing hormone  
 C:Species: Xenopus laevis (African clawed frog)  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 16-Jul-1999  
 C:Accession: 151423  
 R:Hayes, W.P.; Wray, S.; Batley, J.F.  
 Endocrinology 134, 1835-1845, 1994  
 A>Title: The frog GnRH-I gene has a mammalian-like expression pattern and conserved doma  
 A:Reference number: 151423; MUID:94185563  
 A:Accession: 151423  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-89 <HAY>  
 A:Cross-references: GB:L28040; NID:g496291; PIDN:AAA9728.1; PID:g496292  
 C:Genetics:  
 A:Gene: GnRH-I  
 C:Superfamily: gonadoliberin

Query Match 95.2%; Score 60; DB 2; Length 89;  
 Best Local Similarity 90.0%; Pred. No. 0.0022;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EHWSYGLRPG 10

Db 24 QHWSYGLRPG 33  
 :|||||

RESULT 5  
 RHM5G  
 gonadoliberin precursor - mouse  
 N:Alternate names: gonadotropin-releasing hormone (GnRH); luteinizing hormone releas  
 N:Contains: gonadoliberin; gonadoliberin-associated protein (GAP)  
 C:Species: Mus musculus (house mouse)  
 C>Date: 31-Dec-1993 #sequence\_revision 18-Mar-1997 #text\_change 18-Jun-1999  
 C:Accession: A47578  
 R:Mason, A.J.; Hayflick, J.S.; Zoeller, R.T.; Young III, W.S.; Phillips, H.S.; Nikoli  
 Science 234, 1366-1371, 1986  
 A>Title: A deletion truncating the gonadotropin-releasing hormone gene is responsible  
 A:Reference number: A47578; MUID:87069928  
 A:Accession: A47578  
 A:Molecule type: DNA  
 A:Residues: 1-90 <MAS>  
 A:Cross-references: EMBL:M14872; NID:g193576; PIDN:AAA37717.1; PID:g387175  
 C:Genetics:  
 A:Insertions: 45/3; 77/3  
 C:Function:  
 A>Description: gonadoliberin stimulates pituitary secretion of lutropin and follitrop  
 A>Note: gonadoliberin-associated protein may have prolactin release inhibiting activi  
 C:Superfamily: gonadoliberin  
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglyutamic acid  
 F:1-73/Domain: signal sequence #status predicted <SIG>  
 F:22-31/Product: gonadoliberin #status predicted <GAP>  
 F:35-90/Product: gonadoliberin-associated protein #status predicted <GAP>  
 F:32/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic  
 F:31/Modified site: amidated carboxyl end (Gly) (amide in mature form from following

Query Match 95.2%; Score 60; DB 1; Length 90;  
 Best Local Similarity 90.0%; Pred. No. 0.0023;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHWSYGLRPG 10  
 Db 22 QHWSYGLRPG 31

RESULT 6  
 RHNG  
 gonadoliberin precursor [validated] - human  
 N:Alternate names: gonadotropin releasing hormone (GnRH); luteinizing hormone releasi  
 N:Contains: gonadoliberin-associated protein (GAP); progadoliberin  
 C:Species: Homo sapiens (man)  
 C>Date: 17-Mar-1987 #sequence\_revision 21-Jul-1995 #text\_change 08-Dec-2000  
 C:Accession: S05308; A26173; A93342; A90108; A01410; S45718  
 R:Hayflick, J.S.; Adelman, J.P.; Seeburg, P.H.  
 Nucleic Acids Res. 17, 6403-6404, 1989  
 A>Title: The complete nucleotide sequence of the human gonadotropin-releasing hormone  
 A:Reference number: S05308; MUID:89366682  
 A:Accession: S05308  
 A>Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-92 <HAY>  
 A:Cross-references: EMBL:X15215; NID:g31955; PIDN:CAA33285.1; PID:g31956  
 R:Adelman, J.P.; Mason, A.J.; Hayflick, J.S.; Seeburg, P.H.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 179-183, 1986  
 A>Title: Isolation of the gene and hypothalamic cDNA for the common precursor of gona  
 A:Reference number: A94090; MUID:86094338  
 A:Accession: A26173  
 A:Molecule type: mRNA  
 A:Residues: 1-92 <ADE>  
 A:Cross-references: GB:M12578; NID:g183418; PIDN:AAA35916.1; PID:g386749  
 A:Experimental source: hypothalamus  
 R:Seeburg, P.H.; Adelman, J.P.  
 Nature 311, 666-668, 1984  
 A>Title: Characterization of cDNA for precursor of human luteinizing hormone releasin  
 A:Reference number: A93342; MUID:85012739



A:Accession: A93342  
 A:Molecule type: mRNA  
 A:Residues: 1-15, 'S', 17-92 <SEE>  
 A:Cross-references: GB:X01059; NID:g34356; PIDN:CAA25526.1; PID:g34357  
 A:Experimental source: placenta  
 R:Tan, L.; Rousseau, P.  
 A:Title: The chemical identity of the immunoreactive LHRH-like peptide biosynthesized in  
 A:Reference number: A90108; MUID:83126573  
 A:Accession: A90108  
 A:Molecule type: protein  
 A:Residues: 24-33 <TRAN>  
 A:Experimental source: Placental trophoblasts  
 R:Leibovitz, D.; Koch, Y.; Pitzer, F.; Fridkin, M.; Dantes, A.; Baumeister, W.; Amsterda  
 FES Lett. 346, 203-206, 1994  
 A:Title: Sequential degradation of the neuropeptide gonadotropin-releasing hormone by th  
 A:Reference number: S45718; MUID:94283597  
 A:Contents: annotation; degradation pathway of synthetic hormone  
 C:Genetics:  
 A:Gene: GDB:GNRH; LHRH; GRH  
 A:Cross-references: GDB:133746; OMIM:227200; OMIM:152760  
 A:Map position: 8p21-Bp11.2  
 A:Introns: 47/3; 79/3  
 C:Function:  
 A:Description: gonadoliberin stimulates pituitary secretion of lutropin and follitropin  
 A:Note: gonadoliberin-associated protein may have prolactin release inhibiting activity  
 C:Superfamily: gonadoliberin  
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; placenta; pyroglutamic acid  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-92/Product: progonaoliberin #status predicted <PGN>  
 F:24-33/Product: gonadoliberin #status experimental <MAT>  
 F:37-92/Product: gonadoliberin-associated protein #status predicted <GAP>  
 F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experiment  
 F:33/Modified site: amidated carboxyl end (Gly) (amide in mature form from following gly

Query Match 95.2%; Score 60; DB 1; Length 92;  
 Best Local Similarity 90.0%; Pred. No. 0.0023;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||  
 Db 24 QHMSYGLRPG 33

## RESULT 7

RHRTG  
 gonadoliberin precursor - rat  
 N:Alternate names: gonadoliberin-associated protein (GAP); gonadotropin releasing hormo  
 N:Contains: gonadoliberin; prolactin release-inhibiting factor  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 18-Jun-1999  
 C:Accession: A40147; B26173; A48410  
 R:Bond, C.T.; Hayflick, J.S.; Seeburg, P.H.; Adelman, J.P.  
 Mol. Endocrinol. 3, 1257-1262, 1989  
 A:Title: The rat gonadotropin-releasing hormone: SH locus: structure and hypothalamic ex  
 A:Reference number: A40147; MUID:89384661  
 A:Accession: A40147  
 A:Molecule type: DNA  
 A:Residues: 1-92 <BON>  
 A:Cross-references: GB:M31670; NID:g204447; PIDN:AAA1264.1; PID:g204448  
 R:Adelman, J.P.; Mason, A.J.; Hayflick, J.S.; Seeburg, P.H.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 179-183, 1986  
 A:Title: Isolation of the gene and hypothalamic cDNA for the common precursor of gonadot  
 A:Reference number: A94090; MUID:86094358  
 A:Accession: B26173  
 A:Molecule type: mRNA  
 A:Residues: 1-92 <ADE>  
 A:Cross-references: GB:M12579; NID:g204445; PIDN:AAA1263.1; PID:g204446  
 R:Maier, C.C.; Marchetti, B.; Lebeauf, R.D.; Blalock, J.E.  
 Cell. Mol. Neurobiol. 12, 447-454, 1992  
 A:Title: Thymocytes express a mRNA that is identical to hypothalamic luteinizing hormone  
 A:Reference number: A48410; MUID:93105480

A:Accession: A48410  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-92 <MAT>  
 A:Cross-references: GB:S50870; NID:g262059; PIDN:AAB24572.1; PID:g262060  
 A:Experimental source: thymus  
 A:Note: sequence extracted from NCBI backbone (NCBI:121082, NCBI:P.121083)  
 C:Genetics:  
 A:Introns: 47/3; 79/3  
 C:Function:  
 A:Description: stimulates pituitary secretion of lutropin and follitropin  
 A:Note: gonadoliberin-associated protein may have prolactin release inhibiting activi  
 C:Superfamily: gonadoliberin  
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; placenta; pyroglutamic acid  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-92/Product: progonaoliberin #status predicted <PGN>  
 F:24-33/Product: gonadoliberin #status predicted <GPN>  
 F:37-92/Product: prolactin release-inhibiting factor #status predicted <PIF>  
 F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic  
 F:33/Modified site: amidated carboxyl end (Gly) (amide in mature form from following

Query Match 95.2%; Score 60; DB 1; Length 92;  
 Best Local Similarity 90.0%; Pred. No. 0.0023;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||  
 Db 24 QHMSYGLRPG 33

## RESULT 8

RHA01  
 gonadoliberin I - American alligator  
 N:Alternate names: gonadotropin-releasing hormone I  
 C:Species: Alligator mississippiensis (American alligator)  
 C:Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 18-Mar-1997  
 C:Accession: A60066  
 R:Lovejoy, D.A.; Fischer, W.H.; Parker, D.B.; McRory, J.E.; Park, M.; Lance, V.; Swan  
 Regul. Pept. 33, 105-116, 1991  
 A:Title: Primary structure of two forms of gonadotropin-releasing hormone from brains  
 A:Reference number: A60066; MUID:91352338  
 A:Accession: A60066  
 A:Molecule type: protein  
 A:Residues: 1-10 <LOW>  
 C:Superfamily: gonadoliberin  
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental  
 F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 88.9%; Score 56; DB 1; Length 10;  
 Best Local Similarity 80.0%; Pred. No. 0.0011;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||  
 Db 1 QHMSYGLRPG 10

## RESULT 9

gonadoliberin I precursor - chicken  
 N:Alternate names: gonadotropin-releasing hormone I  
 C:Species: Gallus gallus (chicken)  
 C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 16-Jul-1999  
 C:Accession: I50644; S33507  
 R:Dunn, I.C.; Chen, Y.; Hook, C.; Sharp, P.J.; Sang, H.M.  
 J. Mol. Endocrinol. 11, 19-29, 1993  
 A:Title: Characterization of the chicken preprogonadotropin-releasing hormone-I gene  
 A:Reference number: I50644; MUID:94059335  
 A:Accession: I50644  
 A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-92 <DU2>  
 A:Cross-references: EMBL:X69491; NID:9496326; PIDN:CAA49246.1; PID:g311612  
 C:Genetics:  
 A:introns: 47/3; 79/3  
 C:Superfamily: gonadoliberin

Query Match 88.9%; Score 56; DB 2; Length 92;  
 Best Local Similarity 80.0%; Pred. No. 0.011;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 EHMSYGLRPG 10  
 :|||||:  
 DB 24 QHWSYGLQPG 33

## RESULT 10

150739  
 gonadotropin-releasing hormone - Cichlid (Haplochromis burtoni)  
 C:Species: Haplochromis burtoni  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: 150739  
 R:White, S.A.; Kasten, T.L.; Bond, C.T.; Agelman, J.P.; Fernald, R.D.  
 Proc. Natl. Acad. Sci. U.S.A. 92, 8363-8367, 1995  
 A:Title: Three gonadotropin-releasing hormone genes in one organism suggest novel roles  
 A:Reference number: 150739; MUID:95396797  
 A:Accession: 150739  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-98 <WH1>  
 A:Cross-references: EMBL:U31865; NID:9505398; PIDN:MAC59691.1; PID:g905399  
 C:Superfamily: gonadoliberin

Query Match 85.7%; Score 54; DB 2; Length 98;  
 Best Local Similarity 80.0%; Pred. No. 0.026; 1; Indels 0; Gaps 0;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EHMSYGLRPG 10  
 :|||||:  
 DB 23 QHWSYGLSPG 32

## RESULT 11

RH1D5  
 gonadoliberin I precursor - sharpshooth catfish  
 N:Alternate names: gonadoliberin I; gonadoliberin I-associated protein form I; gonadoliberin I-  
 N:Contents: gonadoliberin I-associated protein form I  
 C:Species: Clarias gariepinus (sharpshooth catfish)  
 C:Date: 30-Sep-1993 #sequence\_revision 18-Mar-1997 #text\_change 18-Jun-1999  
 C:Accession: S45602; S45601; JCI242; S42936; S42937  
 R:Boiger, J.; Zandbergen, T.; Andersson, E.; Goos, H.  
 Eur. J. Biochem. 222, 541-549, 1994  
 A:Title: Isolation, characterization and expression of cDNAs encoding the catfish-type 2  
 A:Reference number: S45602; MUID:94291651  
 A:Accession: S45602  
 A:Molecule type: mRNA  
 A:Residues: 1-80 <BOG1>  
 A:Cross-references: EMBL:X78049; NID:9459433; PIDN:CAA54971.1; PID:g459434  
 A:Note: gonadoliberin I-associated protein form I  
 A:Accession: S45601  
 A:Molecule type: mRNA  
 A:Residues: 1-46; S' 48-59; G' 61-80 <BOG2>  
 A:Cross-references: EMBL:X78048; NID:9459431; PIDN:CAA54970.1; PID:g459432  
 A:Note: gonadoliberin I-associated protein form II, presumed to be a polymorphic form  
 R:Boiger, J.; Li, K.W.; Janssen-Dommerholt, C.; Goos, H.  
 Biochem. Biophys. Res. Commun. 187, 127-134, 1992  
 A:Title: Two gonadotropin-releasing hormones from African catfish (Clarias gariepinus).  
 A:Reference number: JCI242; MUID:92392313  
 A:Accession: JCI242  
 A:Molecule type: protein  
 A:Residues: 22-31 <BOG3>

A:Experimental source: brain  
 C:Superfamily: gonadoliberin  
 C:Keywords: amidated carboxyl end; hormone; hypothalamus; pyroglutamic acid  
 F:1-21/Pomatin: signal sequence #status predicted <SIG>  
 F:22-31/Product: gonadoliberin I #status experimental <MAT>  
 F:35-80/Product: gonadoliberin I-associated protein #status predicted <MAT>  
 F:22/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status expert  
 F:31/Modified site: amidated carboxyl end (Gly) (amide in mature form from following

Query Match 79.4%; Score 50; DB 1; Length 80;  
 Best Local Similarity 70.0%; Pred. No. 0.099;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 EHMSYGLRPG 10  
 :|||||:  
 DB 22 QHWSYGLNPG 31

## RESULT 12

JC7393  
 medaka-type gonadotropin-releasing hormone precursor - Japanese medaka  
 C:Species: Oryzias latipes (Japanese medaka)  
 C:Date: 17-Nov-2000 #sequence\_revision 17-Nov-2000 #text\_change 17-Nov-2000  
 C:Accession: JC7393  
 R:Okubo, K.; Amano, M.; Yoshitake, Y.; Sueake, H.; Aida, K.  
 Biochem. Biophys. Res. Commun. 276, 298-303, 2000  
 A:Title: A novel form of gonadotropin-releasing hormone in the medaka, Oryzias latipes  
 A:Reference number: JC7393  
 A:Accession: JC7393  
 A:Contents: Brain  
 A:Molecule type: mRNA  
 A:Residues: 1-91 <OKU>  
 A:Cross-references: DDBJ:AB041333  
 C:Comment: This protein plays the roles as a hypophysiotropic factor, and a physiolog  
 C:Genetics:  
 A:Gene: mdgnrh  
 C:Keywords: brain

Query Match 79.4%; Score 50; DB 2; Length 91;  
 Best Local Similarity 70.0%; Pred. No. 0.11;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 EHMSYGLRPG 10  
 :|||||:  
 DB 22 QHWSYGLSPG 31

## RESULT 13

A21114  
 gonadoliberin - chum salmon  
 C:Species: Oncorhynchus keta (chum salmon)  
 C:Date: 10-Aug-1990 #sequence\_revision 10-Aug-1990 #text\_change 18-Jun-1993  
 C:Accession: A21114  
 R:Sherwood, N.; Elden, L.; Brownstein, M.; Spiess, J.; Rivier, J.; Vale, W.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 2794-2798, 1983  
 A:Title: Characterization of a teleost gonadotropin-releasing hormone.  
 A:Reference number: A21114; MUID:83195140  
 A:Accession: A21114  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 1-10 <SHE>

Query Match 74.6%; Score 47; DB 2; Length 10;  
 Best Local Similarity 70.0%; Pred. No. 0.036;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHMSYGLRPG 10  
 :|||||:  
 DB 1 QHWSYGLWLPG 10

## RESULT 14

151092  
gonadotropin releasing hormone - chinook salmon (fragment)  
C:Species: Oncorhynchus tshawytscha (chinook salmon)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Aug-1997  
C:Accession: I51092  
R:Klungland, H.; Lorens, J.B.; Andersen, O.; Kisen, G.O.; Alestrom, P.  
Mol. Cell. Endocrinol. 84, 167-174, 1992  
A:Title: The Atlantic salmon prepro-gonadotropin releasing hormone gene and mRNA.  
A:Reference number: I51040; M0ID:92267241  
A:Accession: I51092  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-74 <RLU>  
A:Cross-references: EMBL:X79711; NID:g499322; PID:g499323  
C:Genetics:  
A:Gene: GnRH  
A:Introns: 38/3; 65/3

## Query Match

Best Local Similarity 74.6%; Score 47; DB 2; Length 74;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHWSTGLRPG 10

Db 16 QHWSYGLWPG 25

## RESULT 15

151180  
gonadotropin-releasing hormone - cherry salmon  
C:Species: Oncorhynchus masou (cherry salmon)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Aug-1997  
C:Accession: I51180  
R:Suzuki, M.; Hyodo, S.; Kobayashi, M.; Aida, K.; Urano, A.  
J. Mol. Endocrinol. 9, 73-82, 1992  
A:Title: Characterization and localization of mRNA encoding the salmon-type gonadotropin  
A:Reference number: I51180; M0ID:92384893  
A:Accession: I51180  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-82 <SUZ>  
A:Cross-references: GB:S44614; NID:g254824; PID:g254825

## Query Match

Best Local Similarity 74.6%; Score 47; DB 2; Length 82;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHWSTGLRPG 10

Db 24 QHWSYGLWPG 33

Search completed: November 13, 2001, 14:28:18  
Job time: 124 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:27:49 ; Search time 17.9 Seconds

(without alignments)  
19.137 Million cell updates/sec

Title: US-09-412-558-1

Sequence: 1 EHWSYGLRPG 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query length	ID	Description
1	60	95.2	61	1	GON1_SHEEP
2	60	95.2	63	1	GON1_MESAU
3	60	95.2	67	1	GON1_MACMU
4	60	95.2	89	1	GON1_XENIA
5	60	95.2	90	1	GON1_MOUSE
6	60	95.2	91	1	GON1_PIG
7	60	95.2	92	1	GON1_HUMAN
8	60	95.2	92	1	GON1_RAT
9	60	95.2	92	1	GON1_TUPGB
10	56	88.9	10	1	GON1_ALIMI
11	56	88.9	92	1	GON1_CHICK
12	54	85.7	94	1	GON1_HAPBU
13	54	85.7	95	1	GON1_PAGMA
14	54	85.7	95	1	GON1_SPAU
15	51	81.0	92	1	GON1_CAVPO
16	50	79.4	80	1	GON1_CLAGA
17	47	74.6	10	1	GON3_ONCKE
18	47	74.6	74	1	GON3_ONCMY
19	47	74.6	74	1	GON3_ONCTS
20	47	74.6	82	1	GON3_ONOMA
21	47	74.6	82	1	GON3_SALSA
22	47	74.6	89	1	GON3_PORNO
23	47	74.6	89	1	GON3_SALTR
24	47	74.6	90	1	GON3_HAPBU
25	47	74.6	90	1	GON3_PAGMA
26	47	74.6	90	1	GON3_SPAU
27	47	74.6	94	1	GON3_CARAU
28	47	74.6	94	1	GON3_RUTRU
29	42	66.7	10	1	GON2_CHICK
30	42	66.7	10	1	GON2_SQUAC
31	42	66.7	85	1	GON2_HAPBU
32	42	66.7	85	1	GON2_SPAU
33	42	66.7	86	1	GON2_CARAU

34	42	66.7	86	1	GON2_CLAGA	P43306	clarlas gar
35	42	66.7	86	1	GON2_ONCMY	O42241	oncorhynch
36	42	66.7	86	1	GON2_RUMRU	O91330	rullius rut
37	42	66.7	110	1	GON2_SUNMU	O97666	suncus mur
38	42	66.7	110	1	GON2_TUPGB	O95336	tupaia glis
39	42	66.7	120	1	GON2_HUMAN	O43555	homo sapien
40	41	65.1	110	1	YHBJ_ACTAC	P96769	actinobacil
41	41	65.1	315	1	ALD2_MOUSE	P45377	mus musculu
42	40	63.5	551	1	YABN_ECOLI	P33595	escherichia
43	40	63.5	584	1	GUND_CLOCE	P25472	clostridium
44	39	61.9	10	1	GON3_PETMA	P30948	petromyzon
45	38	60.3	432	1	YAIL_RHISN	P55495	rhizobium s

## ALIGNMENTS

RESULT	1	STANDARD:	PRT:	61 AA.
GON1_SHEEP				
ID	GON1_SHEEP			
AC	O28588:			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	PROGONADOLIBERIN I PRECURSOR [CONTAINS: GONADOLIBERIN I (LHRH I)]			
DE	(LUTEINIZING HORMONE RELEASING HORMONE I) (GONADOTROPIN RELEASING HORMONE I) (GNRH I) (LULIBERIN I); GNRH-ASSOCIATED PEPTIDE I)			
DE	(FRAGMENT).			
GN	GNRH1 OR GNRH OR LHRH.			
OS	Ovis aries (Sheep).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Caprinae; Ovis.			
OX	NCBI_TaxID=9940;			
RN	[1]			
RP	SEQUENCE OF 12-61 FROM N.A.			
RC	STRAIN=WESTERN RANGE; TISSUE=Hypothalamus;			
RA	Rodriguez R.E., Wise M.E.;			
RL	Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 1-10.			
RX	MEDLINE=72094314; PubMed=4550508;			
RA	Burgins R., Butcher M., Amoss M., Ling N., Monahan M., Rivier J.,			
RA	Fellows R., Blackwell R., Vale W., Guillemin R.;			
RT	"Primary structure of the ovine hypothalamic luteinizing hormone-releasing factor (LRF) (LH-hypothalamus-LRF-gas chromatography-mass spectrometry-decapeptide-B-man degradation)."			
RT	Proc. Natl. Acad. Sci. U.S.A. 69:278-282(1972).			
CC	-1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS; IT STIMULATES THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING HORMONES.			
CC	-1- SIMILARITY: BELONGS TO THE GNRH FAMILY.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb.slb.ch">license@isb.slb.ch</a> ).			
CC	EMBL: U02517; AAA03433.1; -			
DR	PIR: A93780; RSHSG.			
DR	InterPro: IPR002012; -			
DR	Pfam: PF00446; GNRH; 1.			
DR	PROSITE: PS00473; GNRH; 1.			
KW	Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus; Placenta.			
FT	NON_TER	1	1	
FT	CHAIN	1	>61	PROGONADOLIBERIN I.
FT	PEPTIDE	1	10	GONADOLIBERIN I.
FT	PEPTIDE	14	>61	GNRH-ASSOCIATED PEPTIDE I.
FT	ACT_SITE	3	3	APPEARS TO BE ESSENTIAL FOR BIOLOGICAL

FT MOD\_RES 1 1 ACTIVITY.  
 FT MOD\_RES 10 1 PYROLIDONE CARBOXYLIC ACID.  
 FT NON\_TER 61 61 AMIDATION (G-11 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 61 AA; 6828 MW; 63962A1AE319B8F0 CRC64;

Query Match 95.2%; Score 60; DB 1; Length 61;  
 Best Local Similarity 90.0%; Pred. No. 0.00038;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHMSTGLRPG 10  
 :|||||  
 Db 1 QHMSYGLRPG 10

RESULT 2  
 GONI\_MESAU STANDARD; PRT; 63 AA.  
 ID GONI\_MESAU  
 AC 009163;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PROGNADOLIBERIN I PRECURSOR [CONTAINS: GONADOLIBERIN I (LHRH I)  
 (LUTEINIZING HORMONE RELEASING HORMONE I) (GONADOTROPIN RELEASING  
 HORMONE I) (GNRH I) (LULIBERIN I); GNRH-ASSOCIATED PEPTIDE I]  
 DE (FRAGMENT).  
 GN GNRH1 OR GNRH OR LHRH.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 NC NCBI\_TaxID=10036;  
 RN (1)  
 RP Jansen H.T., Stevens P.J., Zeitler P., Lehman M.N.;  
 RA Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.  
 RL -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS; IT STIMULATES  
 THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING  
 HORMONES.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U91938; AAB51302.1; -  
 DR InterPro: IPR002012; -  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KM Placenta.  
 FT NON\_TER 1 1  
 FT CHAIN 1 >63  
 FT PEPTIDE 1 10  
 FT PEPTIDE 14 >63  
 FT ACT\_SITE 3 3  
 FT MOD\_RES 1 1  
 FT MOD\_RES 10 10  
 FT MOD\_RES 10 10  
 FT NON\_TER 63 63  
 FT SEQUENCE 63 AA; 7370 MW; FC9499567677180 CRC64;

Query Match 95.2%; Score 60; DB 1; Length 63;  
 Best Local Similarity 90.0%; Pred. No. 0.00039;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHMSTGLRPG 10  
 :|||||  
 Db 1 QHMSYGLRPG 10

RESULT 3  
 GONI\_MACMU STANDARD; PRT; 67 AA.  
 ID GONI\_MACMU  
 AC P55247;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PROGNADOLIBERIN I PRECURSOR [CONTAINS: GONADOLIBERIN I (LHRH I)  
 (LUTEINIZING HORMONE RELEASING HORMONE I) (GONADOTROPIN RELEASING  
 HORMONE I) (GNRH I) (LULIBERIN I); GNRH-ASSOCIATED PEPTIDE I]  
 DE (FRAGMENT).  
 GN GNRH1 OR GNRH OR LHRH.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 NC NCBI\_TaxID=9544;  
 RN (1)  
 RP TISSUE-Hypothalamus;  
 RC MEDLINE=95124501; PubMed=7545971;  
 RX Ma Y.J., Costa M.E., Ojeda S.R.;  
 RT "Developmental expression of the genes encoding transforming growth  
 RT factor alpha and its receptor in the hypothalamus of female rhesus  
 RT macaques.";  
 RL Neuroendocrinology 60:346-359(1994).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS; IT STIMULATES  
 THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING  
 HORMONES.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; S75918; AAB33096.1; -  
 DR InterPro: IPR002012; -  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KM Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL 1 5  
 FT CHAIN 6 >67  
 FT PEPTIDE 6 15  
 FT PEPTIDE 19 >67  
 FT ACT\_SITE 8 8  
 FT MOD\_RES 6 6  
 FT MOD\_RES 15 15  
 FT MOD\_RES 15 15  
 FT NON\_TER 67 67  
 FT SEQUENCE 67 AA; 7573 MW; 505394DAA261A3F2 CRC64;

Query Match 95.2%; Score 60; DB 1; Length 67;  
 Best Local Similarity 90.0%; Pred. No. 0.00042;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHMSTGLRPG 10  
 :|||||

```

DE  DE  HORMONE ) (GNRH I) (LULIBERIN I); PROLACTIN RELEASE-INHIBITING FACTOR
DE  I].
GN  GNRH1 OR GNRH.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=87065928; PubMed=3024317;
RA  Mason A.J., Hayflick J.S., Zoeller R.T., Young W.S. III,
RA  Phillips H.S., Nikolic K., Seeburg P.H.;
RT  "A deletion truncating the gonadotropin-releasing hormone gene is
RT  responsible for hypogonadism in the hpg mouse."
RL  Science 234:1366-1371(1986).
CC  -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS; IT STIMULATES
CC  THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING
CC  HORMONES.
CC  -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
CC  -----
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isdb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL; M14872; AAA37717.1; -.
DR  MGD; MGI:95789; GnRH.
DR  InterPro; IPR002012; -.
DR  Pfam; PF00446; GNRH; 1.
DR  PROSITE; PS00473; GNRH; 1.
KW  Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW  Placenta; Signal.
FT  SIGNAL 1 21
FT  CHAIN 22 90 PROGNADOLIBERIN I.
FT  PEPTIDE 22 31 GONADOLIBERIN I.
FT  PEPTIDE 35 90 PROLACTIN RELEASE-INHIBITING FACTOR I.
FT  ACT_SITE 24 24 APPEARS TO BE ESSENTIAL FOR BIOLOGICAL
FT  ACTIVITY.
FT  MOD_RES 22 22 PYRROLIDONE CARBOXYLIC ACID.
FT  MOD_RES 31 31 AMIDATION (G-32 PROVIDE AMIDE GROUP).
SQ  SEQUENCE 90 AA; 10337 MW; 1C0766FA4826E4D9 CRC64;

Query Match 95.2%; Score 60; DB 1; Length 90;
Best Local Similarity 90.0%; Pred. NO. 0.00056;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EHMSTGLRPG 10
   :|11111111
Db 22 QHWSYGLRPG 31

RESULT 6
GONL_PIG STANDARD; PRT; 91 AA.
AC P49921;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROGNADOLIBERIN I PRECURSOR [CONTAINS: GONADOLIBERIN I (LHRH I)
DE (LUTEINIZING HORMONE RELEASING HORMONE I) (GONADOTROPIN RELEASING
DE HORMONE I) (GNRH I) (LULIBERIN I); GNRH-ASSOCIATED PEPTIDE I].
GN GNRH1 OR GNRH.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Hypothalamus;

```

RA Weesner G.D., Matteri R.L., Becker B.A.;  
 RL Submitted (MAY-1994) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 24-33.  
 RX MEDLINE=72114303; PubMed=4946067;  
 RA Baba Y., Matsuo H., Schally A.V.;  
 RT "Structure of the porcine LH- and FSH-releasing hormone. II.  
 RT Confirmation of the proposed structure by conventional sequential  
 RT analyses.";  
 RL Biochem. Biophys. Res. Commun. 44:459-463(1971).  
 RN [3]  
 RP SYNTHESIS OF GONADOLIBERIN.  
 RX MEDLINE=72065376; PubMed=4942726;  
 RA Matsuo H., Arimura A., Nair R.M.G., Schally A.V.;  
 RT "Synthesis of the porcine LH- and FSH-releasing hormone by the solid-  
 RT phase method.";  
 RL Biochem. Biophys. Res. Commun. 45:822-827(1971).  
 RN [4]  
 RP SYNTHESIS OF GONADOLIBERIN.  
 RX MEDLINE=72117544; PubMed=4946275;  
 RA Baba Y., Arimura A., Schally A.V.;  
 RT "On the tryptophan residue in porcine LH and FSH-releasing hormone.";  
 RL Biochem. Biophys. Res. Commun. 45:483-487(1971).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS; IT STIMULATES  
 CC THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING  
 CC HORMONES.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: J32664; AAA31066.1; -  
 DR PIR: A01411; RHPGG.  
 DR InterPro: IPR002012; -  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KM Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KW Placenta; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 91 PROGONADOLIBERIN I.  
 FT PEPTIDE 24 33 GONADOLIBERIN I.  
 FT PEPTIDE 34 91 GNRH-ASSOCIATED PEPTIDE I.  
 FT ACT\_SITE 26 26 APPEARS TO BE ESSENTIAL FOR BIOLOGICAL  
 FT ACTIVITY.  
 FT MOD\_RES 24 24 PYRROLIDONE CARBOXYLIC ACID  
 FT MOD\_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP).  
 FT SEQUENCE 91 AA; 10090 MW; 8340474F32DDA99 CRC64;  
 SQ  
 Query Match 95.2%; Score 60; DB 1; Length 91;  
 Best Local Similarity 90.0%; Pred. No. 0.00057;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHWSTGLRPG 10  
 DB 24 QHWSYGLRPG 33  
 RESULT 7  
 GONI\_HUMAN STANDARD; PRT; 92 AA.  
 AC P01148; 1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-APR-1988 (Rel. 07, Last annotation update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROGONADOLIBERIN I PRECURSOR [CONTAINS: GONADOLIBERIN I (LHRH I)  
 DE (LUTEINIZING HORMONE RELEASING HORMONE I) (GONADOTROPIN RELEASING  
 DE HORMONE I) (GNRH I) (LULIBERIN I) (GONADORELIN); GNRH-ASSOCIATED

DE PEPTIDE I1.  
 GN GNRH OR LHRH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89366682; PubMed=2671939;  
 RA Hayflick J.S., Adelman J.P., Seeburg P.H.;  
 RT "The complete nucleotide sequence of the human gonadotropin-releasing  
 RT hormone gene.";  
 RL Nucleic Acids Res. 17:6403-6403(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86094338; PubMed=2867548;  
 RA Adelman J.P., Mason A.J., Hayflick J.S., Seeburg P.H.;  
 RT "Isolation of the gene and hypothalamic cDNA for the common precursor  
 RT of gonadotropin-releasing hormone and prolactin release-inhibiting  
 RT factor in human and rat.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:179-183(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85012739; PubMed=6090951;  
 RA Seeburg P.H., Adelman J.P.;  
 RT "Characterization of cDNA for precursor of human luteinizing hormone  
 RT releasing hormone.";  
 RL Nature 311:666-668(1984).  
 RN [4]  
 RP SEQUENCE OF 24-33.  
 RX MEDLINE=83126573; PubMed=6760865;  
 RA Tan L., Kousseau P.;  
 RT "The chemical identity of the immunoreactive LHRH-like peptide  
 RT biosynthesized in the human placenta.";  
 RL Biochem. Biophys. Res. Commun. 109:1061-1071(1982).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS; IT STIMULATES  
 CC THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING  
 CC HORMONES.  
 CC -1- PHARMACEUTICAL: AVAILABLE UNDER THE NAMES FACTREL (AYERST LABS),  
 CC LUTREPLIS OR LUTRELEF (FERRING PHARMACEUTICALS) AND RELISORM  
 CC (SERONO).  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X01059; CAA25526.1; -  
 DR EMBL: M12578; AAA35916.1; -  
 DR EMBL: X15215; CAA33285.1; -  
 DR PIR: A01410; RHHUG.  
 DR PIR: A26173; A26173.  
 DR PIR: S05308; S05308.  
 DR MIM: 152760; -  
 DR InterPro: IPR002012; -  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KM Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KW Placenta; Pharmaceutical; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 92 PROGONADOLIBERIN I.  
 FT PEPTIDE 24 33 GONADOLIBERIN I.  
 FT PEPTIDE 37 92 GNRH-ASSOCIATED PEPTIDE I.  
 FT ACT\_SITE 26 26 APPEARS TO BE ESSENTIAL FOR BIOLOGICAL  
 FT ACTIVITY.  
 FT MOD\_RES 24 24 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP).  
 FT CONFLICT 16 16 W -> S (IN REF. 3).  
 FT SEQUENCE 92 AA; 10380 MW; 30A72221B076FA79 CRC64;



Query Match 95.2% Score 60; DB 1; Length 92;  
 Best Local Similarity 90.0%; Pred. No. 0.00057;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EHWSYGLRPG 10  
 :|||||  
 Db 24 QHWSYGLRPG 33

RESULT 8

CON1_RAT	STANDARD	PRT	92 AA
ID	P07490		
AC	01-APR-1988 (Rel. 07, Created)		
DT	01-APR-1988 (Rel. 07, Last sequence update)		
DT	01-OCT-2000 (Rel. 40, Last annotation update)		
DE	PROGONADOLIBERIN I PRECURSOR [CONTAINS: GONADOLIBERIN I (LHRH I)]		
DE	(LUTEINIZING HORMONE RELEASING HORMONE I) (GONADOTROPIN RELEASING HORMONE I) (GNRH I) (LULIBERIN I); PROLACTIN RELEASE-INHIBITING FACTOR I]		
GN	GNRH1 OR GNRH.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86094338; PubMed=2867548;		
RA	Adelman J.P., Mason A.J., Hayflick J.S., Seeburg P.H.;		
RT	*Isolation of the gene and hypothalamic cDNA for the common precursor of gonadotropin-releasing hormone and prolactin release-inhibiting factor in human and rat.*;		
RT	Proc. Natl. Acad. Sci. U.S.A. 83:179-183(1986).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RX	MEDLINE=89384661; PubMed=2476669;		
RA	Bond C.T., Hayflick J.S., Seeburg P.H., Adelman J.P.;		
RT	*The rat gonadotropin-releasing hormone SH locus: structure and hypothalamic expression.*;		
RT	Mol. Endocrinol. 3:1257-1262(1989).		
RL	[3]		
RN	SEQUENCE FROM N.A.		
RP	TISSUE=Thymus;		
RX	MEDLINE=93105480; PubMed=1468115;		
RA	Maler C.C., Marchetti B., Leboeuf R.D., Blalock J.E.;		
RT	*Thymocytes express a mRNA that is identical to hypothalamic luteinizing hormone-releasing hormone mRNA.*;		
RT	Cell. Mol. Neurobiol. 12:447-454(1992).		
RL	[4]		
RN	SEQUENCE OF 1-47 FROM N.A.		
RP	TISSUE=Heart;		
RX	MEDLINE=87149087; PubMed=3547652;		
RA	Adelman J.P., Bond C.T., Douglass J., Herbert E.;		
RT	*Two mammalian genes transcribed from opposite strands of the same DNA locus.*;		
RT	Science 235:1514-1517(1987).		
RL	[5]		
CC	-1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS. IT STIMULATES THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING HORMONES.		
CC	-1- TISSUE SPECIFICITY: CENTRAL NERVOUS SYSTEM.		
CC	-1- SIMILARITY: BELONGS TO THE GNRH FAMILY.		
CC	-----		
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
CC	EMBL: S50870; AAB24572.1; .		

DR EMBL: M12579; AAA41263.1; .  
 DR EMBL: M31670; AAA41264.1; .  
 DR EMBL: M15527; AAA42141.1; ALT\_SEQ.  
 DR EMBL: M15528; AAA42139.1; .  
 DR EMBL: M15528; .; NOT\_ANNOTATED\_CDS.  
 DR PIR: B26173; RHRTG.  
 DR PIR: A48410; A48410.  
 DR InterPro: IPR002012; .  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KW Placenta; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 92  
 FT PEPTIDE 24 33  
 FT PEPTIDE 37 92  
 FT ACT\_SITE 26 26  
 FT MOD\_RES 24 24  
 FT MOD\_RES 33 33  
 SQ SEQUENCE 92 AA; 10500 MW; 494B5C64DA8A5EB3 CRC64;

Query Match 95.2% Score 60; DB 1; Length 92;  
 Best Local Similarity 90.0%; Pred. No. 0.00057;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 EHWSYGLRPG 10  
 :|||||  
 Db 24 QHWSYGLRPG 33

RESULT 9

CON1_TUPGB	STANDARD	PRT	92 AA
ID	G095335;		
AC	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	PROGONADOLIBERIN I PRECURSOR [CONTAINS: GONADOLIBERIN I (LHRH I)]		
DE	(LUTEINIZING HORMONE RELEASING HORMONE I) (GONADOTROPIN RELEASING HORMONE I) (GNRH I) (LULIBERIN I); GNRH-ASSOCIATED PEPTIDE I].		
DE	GNRH1 OR GNRH.		
GN	Tupaia glis belangeri (Common tree shrew).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Scandentia; Tupaiidae; Tupai.		
OX	NCBI_TaxID=9396;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RP	TISSUE=Hypothalamus;		
RX	MEDLINE=97079639; PubMed=8921350;		
RA	Rasten T.L., White S.A., Norton T.T., Bond C.T., Adelman J.P.;		
RT	*Characterization of two new preproGNRH mRNAs in the tree shrew: first direct evidence for mesencephalic GNRH gene expression in a placental mammal.*;		
RT	Gen. Comp. Endocrinol. 104:7-19(1996).		
RL	[2]		
CC	-1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS. IT STIMULATES THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING HORMONES.		
CC	-1- SIMILARITY: BELONGS TO THE GNRH FAMILY.		
CC	-----		
CC	-1- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
CC	EMBL: U63326; AAB16837.1; .		
CC	InterPro: IPR002012; .		
CC	Pfam: PF00446; GNRH; 1.		

DR PROSITE: PS00473; GNRH: 1.  
 KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KW placenta; Signal.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT CHAIN 24 92 PROGNADOLIBERIN I.  
 FT PEPTIDE 24 33 GNRH-ASSOCIATED PEPTIDE I.  
 FT PEPTIDE 37 92 GNRH-ASSOCIATED PEPTIDE I.  
 FT ACT\_SITE 26 26 APPEARS TO BE ESSENTIAL FOR BIOLOGICAL  
 FT ACTIVITY.  
 FT MOD\_RES 24 24 PYRROLIDONE CARBOXYLIC ACID (BY  
 FT MOD\_RES 33 33 SIMILARITY).  
 FT MOD\_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP) (BY  
 FT STIMILARITY).  
 SQ SEQUENCE 92 AA; 10197 MW; 4FDBF2C58CF5F63B CRC64;

Query Match 95.2%; Score 60; DB 1; Length 92;  
 Best Local Similarity 90.0%; Pred. No. 0.00057;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||  
 DB 24 QHWSYGLRPG 33

RESULT 10  
 GONI\_ALMT STANDARD; PRT; 10 AA.  
 ID GONI\_ALMT  
 AC P37041; P20407;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GONADOLIBERIN I (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I) (LH-RH I)  
 DE (LUTERININ I)  
 OS Alligator mississippiensis (American alligator).  
 OS Alligator mississippiensis (American alligator).  
 OC Archosauria; Crocodylidae; Alligatorine; Alligator.  
 OC NCBI\_TaxID=8496;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=91552338; PubMed=1882082;  
 RA Lovejoy D.A., Fischer W.H., Parker D.B., McRoy J.E., Park M.,  
 RA Lance V., Swanson P., Rivier J.E., Sherwood N.M.;  
 RT Primary structure of two forms of gonadotropin-releasing hormone  
 RT from brains of the American alligator (Alligator mississippiensis).";  
 RL Regul. Pept. 33:105-116(1991).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 CC PIR: A60066; RHA01.  
 DR InterPro: IPR002012; -  
 DR Pfam: PF00446; GNRH: 1.  
 DR PROSITE: PS00473; GNRH: 1.  
 KW Hormone; Amidation; Hypothalamus.  
 FT MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 10 10 AMIDATION.  
 SQ SEQUENCE 10 AA; 1172 MW; 284B3D7286B45A3 CRC64;

Query Match 88.9%; Score 56; DB 1; Length 10;  
 Best Local Similarity 80.0%; Pred. No. 0.00032;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||  
 DB 1 QHWSYGLRPG 10

RESULT 11  
 GONI\_CHICK STANDARD; PRT; 92 AA.  
 ID GONI\_CHICK  
 AC P37042; P20407;  
 DT 01-FEB-1991 (Rel. 17, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE PROGNADOLIBERIN I PRECURSOR (CONTAINS: GONADOLIBERIN I (LH-RH I)  
 DE (LUTERINIZING HORMONE RELEASING HORMONE I) (GONADOTROPIN RELEASING  
 DE HORMONE I) (GNRH I) (LUTERIN I) (GNRH-ASSOCIATED PEPTIDE I).  
 OS Gallus gallus (Chicken).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=WHITE LEGHORN;  
 RX MEDLINE=94059355; PubMed=7902095;  
 RA Dunn I.C., Chen Y., Hook C., Sharp P.J., Sang H.M.;  
 RT "Characterization of the chicken pregonadotropin-releasing  
 RT hormone-I gene.";  
 RL J. Mol. Endocrinol. 11:19-29(1993).  
 RN [2]  
 RP SEQUENCE OF 24-33.  
 RC TISSUE=Hypothalamus;  
 RX MEDLINE=82265778; PubMed=7050119;  
 RA King J.A., Millar R.P.;  
 RT "Structure of chicken hypothalamic luteinizing hormone-releasing  
 RT hormone. II. Isolation and characterization.";  
 RL J. Biol. Chem. 257:10729-10732(1982).  
 RN [3]  
 RP SEQUENCE OF 24-33.  
 RC TISSUE=Hypothalamus;  
 RA King J.A., Millar R.P.;  
 RT "Structure of avian hypothalamic gonadotropin-releasing hormone.";  
 RL S. Afr. J. Sci. 78:124-125(1982).  
 RN [4]  
 RP SYNTHESIS OF 24-33.  
 RX MEDLINE=82265777; PubMed=7050118;  
 RA King J.A., Millar R.P.;  
 RT "Structure of chicken hypothalamic luteinizing hormone-releasing  
 RT hormone. I. Structural determination on partially purified  
 RT material.";  
 RL J. Biol. Chem. 257:10722-10728(1982).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X69491; CAA49246.1; -  
 CC DR PIR: S33507; S33507.  
 DR InterPro: IPR002012; -  
 DR Pfam: PF00446; GNRH: 1.  
 DR PROSITE: PS00473; GNRH: 1.  
 KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;  
 KW Signal.  
 FT SIGNAL 1 23 PROGNADOLIBERIN I.  
 FT CHAIN 24 92 GONADOLIBERIN I.  
 FT PEPTIDE 24 33 GNRH-ASSOCIATED PEPTIDE I.  
 FT PEPTIDE 37 92 GNRH-ASSOCIATED PEPTIDE I.  
 FT MOD\_RES 24 24 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 92 AA; 10206 MW; 61AEB7EBAF508B6A CRC64;

Query Match 88.9%; Score 56; DB 1; Length 92;  
 Best Local Similarity 80.0%; Pred. No. 0.0029;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||

DB	24	OHMSYGLD.PG	33
RESULT	12		
GN1_HAPBU		STANDARD:	PRT: 94 AA.
AC	P51918: O93387:		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	GNADOLIBERIN I PRECURSOR (GNADOTROPIN-RELEASING HORMONE I) (GNRH-I) (LH-RH I) (LULIBERIN I).		
GN	GNRH1.		
OS	Haplochromis burtoni.		
OC	Actinoptera: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:		
OC	Acanthopterygii, Neopterygii, Teleostei, Euteleostei, Neoteleostei:		
OC	Acanthomorpha: Acanthopterygii: Percomorpha: Perciformes: Labroidae:		
OC	Cichlidae: Astatotilapia.		
OX	NCBI.Taxid=8153:		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=95396797: Pubmed=7667296:		
RX	White S.A., Kassten T.L., Bond C.T., Adelman J.P., Fernald R.D.:		
RA	White R.B., Fernald R.D.:		
RT	"Ontogeny of gonadotropin-releasing hormone (GNRH) gene expression		
RT	reveals a distinct origin for GNRH-containing neurons in the		
RT	midbrain."		
RL	Proc. Natl. Acad. Sci. U.S.A. 92:8363-8367(1995).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=99061842: Pubmed=9843638:		
RX	White R.B., Fernald R.D.:		
RA	White S.A., Francis R.C., Fernald R.D., Licht P., Marby C.,		
RT	Sherwood N.M.:		
RT	"Primary structure of solitary form of gonadotropin-releasing hormone		
RT	(GNRH) in cichlid pituitary: three forms of GNRH in brain of cichlid		
RT	and pumpkinseed fish."		
RL	Regul. Pept. 57:43-53(1995).		
CC	-1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS. MAY BE		
CC	RESPONSIBLE FOR THE REGULATION OF THE HYPOTHALAMIC-PITUITARY-		
CC	GNADAL AXIS.		
CC	-1- TISSUE SPECIFICITY: SYNTHESIZED IN PROOPTIC NEURONS AND IS		
CC	TRANSPORTED TO THE PITUITARY IN THE PROOPTIC-HYPHYSICAL AXONS.		
CC	-1- MASS SPECTROMETRY: MW=1113.9; METHOD=MALDI; RANGE=23-32.		
CC	-1- SIMILARITY: BELONGS TO THE GNRH FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
CC	EMBL: U31865: AAC59691.1: -		
DR	EMBL: AF076961: AAC27716.1: -		
DR	InterPro: IPR002012: -		
DR	Pfam: PF00446: GNRH.1.		
DR	PROSITE: PS00473: GNRH.1.		
KW	Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;		
KW	Signal: Multigene family.		
FT	SIGNAL	1	22
FT	CHAR1	23	94
FT	PEPTIDE	23	32
FT	PEPTIDE	36	94
FT	MOD_RES	23	23
			PROGNADOLIBERIN I.
			GNADOLIBERIN I.
			GNRH-ASSOCIATED PEPTIDE I (POTENTIAL).
			PYRROLIDONE CARBOXYLIC ACID.

FT	MOD_RES	32	32	AMIDATION (G-33 PROVIDE AMIDE GROUP).
FT	CONFLICT	86	94	ENGHRTFAK -> KMDTGHSRNERFL (IN REF. 1)
SQ	SEQUENCE	94 AA;	10382 MW;	E57DBA83333278D7 CXC64;

Query Match	85.78;	Score 54;	DB 1;	Length 94;
Best Local Similarity	80.08;	Pred. No. 0.0066;		
Matches	8;	Conservative	1;	Mismatches 1;
				Indels 0;
				Gaps 0

QY	1	EHWSYGLRPG	10
		:	
Db	23	QHWSYGLSPG	32

RESULT	13	
GONI_PAGMA		
ID	GONI_PAGMA	STANDARD;
		PRT;
		95 AA

```

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GONADOLIBERIN I PRECURSOR (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I)
DE (LH-RH I) (LULIBERIN I).
OS Pagrus major (Red sea bream) (Chrysophrys major).
OC Actinopterygii, Neopterygii, Teleostei, Euteleostei, Neoteleostei,
OC Acanthomorphi, Acanthopterygii, Percomorpha, Perciformes, Percoidae;
OC Sparidae; Chrysophrys.
OX NCBI_TaxID=8171;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Brain:
RA Okuzawa K., Granneman J., Bogerd J., Goos H., Zohar Y., Kagawa H.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY
CC SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE GNRH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; D86582; BAA13129.1; -.
CC InterPro: IPR002012; -.
CC Pfam: PF00446; GNRH; 1.
CC PROSITE: PS00473; GNRH; 1.
DR Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
KW Signal; Multigene family.
FT SIGNAL 1 23
FT CHAIN 24 95
FT PEPTIDE 24 33
FT PEPTIDE 24 33
FT MOD_RES 24 24
FT MOD_RES 24 24
FT MOD_RES 33 33
FT MOD_RES 33 33
SQ SEQUENCE 95 AA: 10566 MW:
SQ 61E79C990328D73E CRC64;

Query Match 85.7%; Score 54; DB 1; Length 95;
Best Local Similarity 80.0%; Pred. No. 0.0067;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10
:||||| 11
DB 24 QHMSTGLSPG 33

```

```

GNL SPAAU STANDARD: PRT: 95 AA.
ID GNL SPAAU
AC P51919;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GONADOLIBERIN I PRECURSOR (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I)
DE (LH-RH I) (LULIBERIN I) (SHGNRH).
OS Sparus aurata (gilthead sea bream).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Percoidae;
OC Sparidae; Sparus.
OC NCBI_TaxID=8175;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95268499; PubMed-7749463;
RA Gotthelf Y., Elizur A., Chow M., Chen T.T., Zohar Y.;
RT "Molecular cloning and characterization of a novel gonadotropin-
  releasing hormone from the gilthead seabream (Sparus aurata).";
  Mol. Mar. Biol. Biotechnol. 4:27-35(1995).
RN [2]
RP SEQUENCE OF 26-35.
RC TISSUE=Brain;
RX MEDLINE=95083645; PubMed-7991588;
RA Powell J.F., Zohar Y., Elizur A., Park M., Fischer W.H.,
  Craig A.G., Rivier J.E., Lovejoy D.A., Sherwood N.M.;
RT "Three forms of gonadotropin-releasing hormone characterized from
  brains of one species.";
  Proc. Natl. Acad. Sci. U.S.A. 91:12081-12085(1994).
CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.
CC -1- MASS SPECTROMETRY: MW=1113.6; METHOD=MALDI; RANGE=26-35.
CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
CC EMBL: U30320; AAA75469.1; -.
DR InterPro: IPR002012; -.
DR Pfam: PF00446; GNRH; 1.
DR PROSITE: PS00473; GNRH; 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
  Signal; Multigene family.
FT SIGNAL 1 25
FT CHAIN 26 95 PROGNADOLIBERIN I.
FT PEPTIDE 26 95 GONADOLIBERIN I.
FT PEPTIDE 39 95 GNRH-ASSOCIATED PEPTIDE I (POTENTIAL).
FT MOD_RES 26 26 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 35 35 AMIDATION (G-36 PROVIDE AMIDE GROUP).
SQ SEQUENCE 95 AA; 10753 MW; 49313FD6B87DA CRC64;

Query Match 85.7%; Score 54; DB 1; Length 95;
Best Local Similarity 80.0%; Pred. No. 0.0067;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
DB 26 QHWSYGLSPG 35

RESULT 15
GNL CAVPO STANDARD: PRT: 92 AA.
ID GNL CAVPO
AC 054713;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

```

```

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROGNADOLIBERIN I PRECURSOR (CONTAINS: GONADOLIBERIN I (LHRH I)
DE (LUTEINIZING HORMONE RELEASING HORMONE I) (GONADOTROPIN RELEASING
DE HORMONE I) (GNRH I) (LULIBERIN I); GNRH-ASSOCIATED PEPTIDE I).
OS GNRH I OR GNRH OR LHRH.
GN Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Rodentia; Hystriognath; Cavidae; Cavia.
OC NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY WHITE; TISSUE=Hypothalamus;
RX MEDLINE=97462693; PubMed-932920;
RA Jimenez-Linan M., Rubin B.S., King J.C.;
RT "Examination of guinea pig luteinizing hormone-releasing hormone gene
  reveals a unique decapeptide and existence of two transcripts in the
  brain.";
  Endocrinology 138:4123-4130(1997).
CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS; IT STIMULATES
  THE SECRETION OF BOTH LUTEINIZING AND FOLLICLE-STIMULATING
  HORMONES.
CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
CC EMBL: AF033346; AAB87688.1; -.
DR InterPro: IPR002012; -.
DR PROSITE: PS00473; GNRH; 1.
KW Cleavage on pair of basic residues; Hormone; Amidation; Hypothalamus;
  Placenta; Signal.
FT SIGNAL 1 23
FT CHAIN 24 92 PROGNADOLIBERIN I.
FT PEPTIDE 24 92 GONADOLIBERIN I.
FT PEPTIDE 37 92 GNRH-ASSOCIATED PEPTIDE I.
FT ACT_SITE 26 26 APPEARS TO BE ESSENTIAL FOR BIOLOGICAL
  ACTIVITY (BY SIMILARITY).
FT MOD_RES 24 24 PYRROLIDONE CARBOXYLIC ACID (BY
  SIMILARITY).
FT MOD_RES 33 33 AMIDATION (G-34 PROVIDE AMIDE GROUP) (BY
  SIMILARITY).
FT MOD_RES 33 33 ACETYLATION (G-34 PROVIDE AMIDE GROUP) (BY
  SIMILARITY).
SQ SEQUENCE 92 AA; 10279 MW; AC74613F45D663 CRC64;

Query Match 81.0%; Score 51; DB 1; Length 92;
Best Local Similarity 70.0%; Pred. No. 0.022;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
DB 24 QYWSYGLSPG 33

Search completed: November 13, 2001, 14:34:24
Job time: 395 sec

```



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2001, 14:26:54 ; Search time 46.47 seconds  
(without alignments)  
28.471 Million cell updates/sec

Title: US-09-412-558-1  
Perfect score: 63  
Sequence: 1 EHMSYGLRPG 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mhc:\*  
8: SP-organelle:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-unclassified:\*  
13: SP-vertebrate:\*  
14: SP-virus:\*

Prod. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	95.2	91	Q9PRH0	Q9PRH0 anguilla ja
2	54	85.7	92	Q9YI26	Q9YI26 sparus aura
3	54	85.7	95	Q73812	Q73812 morone saxa
4	54	85.7	99	Q91A10	Q91A10 dicentrarch
5	52	82.5	90	Q91A02	Q91A02 rana dybows
6	50	79.4	91	Q9DGC8	Q9DGC8 oryzias lat
7	49	77.8	10	P81749	P81749 clupea pall
8	47	74.6	10	Q9W751	Q9W751 clupea pall
9	47	74.6	33	Q9W750	Q9W750 oncorhynch
10	47	74.6	33	Q9PT34	Q9PT34 oncorhynch
11	47	74.6	82	Q92094	Q92094 oncorhynch
12	47	74.6	82	Q9W7G1	Q9W7G1 oncorhynch
13	47	74.6	82	Q91800	Q91800 oncorhynch
14	47	74.6	82	Q918P9	Q918P9 oncorhynch
15	47	74.6	88	Q9PSY9	Q9PSY9 sparus aura
16	47	74.6	90	Q91A09	Q91A09 dicentrarch
17	47	74.6	90	Q9DD49	Q9DD49 oryzias lat
18	47	74.6	94	Q9DEH6	Q9DEH6 carassius a
19	47	74.6	94	Q9DEH5	Q9DEH5 carassius a

20	47	74.6	94	Q9DDH8	Q9DDH8 brachydanio
21	42	66.7	10	P81750	P81750 clupea pall
22	42	66.7	75	Q9RTV0	Q9RTV0 trichosurus
23	42	66.7	80	Q9DGC9	Q9DGC9 oryzias lat
24	42	66.7	85	Q73811	Q73811 morone saxa
25	42	66.7	85	Q91A08	Q91A08 dicentrarch
26	42	66.7	86	Q42471	Q42471 carassius a
27	42	66.7	86	Q9PW69	Q9PW69 typhlonecte
28	42	66.7	86	Q9PT25	Q9PT25 oncorhynch
29	42	66.7	87	Q9PR13	Q9PR13 anguilla ja
30	42	66.7	93	Q9DG36	Q9DG36 rana catesb
31	42	66.7	107	Q9T513	Q9T513 macaca mula
32	42	66.7	114	Q9T655	Q9T655 macaca mula
33	42	66.7	868	Q9HER3	Q9HER3 homo sapien
34	42	66.7	828	Q9S2R5	Q9S2R5 arabidopsis
35	41	65.1	315	P91045	P91045 caenorhabdi
36	41	65.1	316	Q08782	Q08782 cricetus
37	41	65.1	532	Q04866	Q04866 caenorhabdi
38	40	63.5	283	Q9F2U4	Q9F2U4 streptomyce
39	40	63.5	565	Q9KM11	Q9KM11 vibrio chol
40	40	63.5	1000	Q9PCD0	Q9PCD0 xyella fas
41	39	61.9	101	Q79746	Q79746 omanosaura
42	39	61.9	322	Q9F3C9	Q9F3C9 streptomyce
43	39	61.9	379	Q9LV12	Q9LV12 arabidopsis
44	39	61.9	388	Q9YD14	Q9YD14 aeropyrum p
45	39	61.9	417	Q9TX12	Q9TX12 caenorhabdi

## ALIGNMENTS

RESULT 1  
ID Q9PRH0 PRELIMINARY: PRT: 91 AA.  
AC Q9PRH0;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE GONADOTROPIN PRECURSOR(GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
DE (LULIBERIN).  
OS Anguilla japonica (Japanese eel).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;  
OC Anguillidae; Anguilla.  
OX NCBI\_TaxID=7937;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RA Okubo K., Suetake H., Aida K.;  
RT "Expression of two gonadotropin-releasing hormone (GNRH) precursor  
genes in various tissues of the Japanese eel and evolution of GNRH.",  
RL Zool. Sci. 16:471-478(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Okubo K., Suetake H., Aida K.;  
RT "A splicing variant for the prepro-mammalian gonadotropin-releasing  
hormone (prepro-mGNRH) mRNA is present in the brain and various  
peripheral tissues of the Japanese eel.",  
RL Zool. Sci. 16:645-651(1999).  
CC - FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY  
SIMILARITY).  
CC - SIMILARITY: TO THE GNRH FAMILY.  
CC EMBL: AB026989; BAA82608.1; -;  
DR EMBL: AB026991; BAA83597.1; -;  
DR InterPro: IPR002012; -;  
DR Pfam: PF00446; GNRH; 1.  
DR PROSITE: PS00473; GNRH; 1.  
KW Amidation; Hormone; Signal.  
FT SIGNAL. 1 22  
FT CHAIN 23 91 GNRH  
FT CHAIN 33 91 GNRH ASSOCIATED PEPTIDE.  
SQ SEQUENCE 91 AA: 9893 MW; BA15C9DC08434A7B CRC64;

Query Match 95.2%; Score 60; DB 13; Length 91;  
 Best Local Similarity 90.0%; Pred. No. 0.0029;  
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||  
 Db 23 QHMSYGLRPG 32

## RESULT 2

OY126 PRELIMINARY; PRT; 87 AA.

AC OY126; 01-MAY-1999 (TREMUREL. 10, Created)  
 DT 01-MAY-1999 (TREMUREL. 10, last sequence update)  
 DT 01-MAR-2001 (TREMUREL. 16, last annotation update)  
 DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LULIBERIN) (FRAGMENT).  
 OS Sparus aurata (Gilthead sea bream).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;  
 OC Sparidae; Sparus.  
 OX NCBI\_TaxID=8175;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=OVARY;  
 RA Nabissi M.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 CC EMBL: AF046801; AAD02427.1; -.  
 DR InterPro: IPR002012; -.  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Amidation; Hormone.  
 FT NON\_TER 1 1  
 FT 87 87  
 FT SEQUENCE 87 AA; 9871 MW; 0D246353D96782A CRC64;

Query Match 85.7%; Score 54; DB 13; Length 87;  
 Best Local Similarity 80.0%; Pred. No. 0.031; 1; Indels 0; Gaps 0;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||  
 Db 21 QHMSYGLRPG 30

RESULT 3  
 OY126 PRELIMINARY; PRT; 95 AA.

AC OY126; 01-AUG-1998 (TREMUREL. 07, Created)  
 DT 01-AUG-1998 (TREMUREL. 07, last sequence update)  
 DT 01-MAR-2001 (TREMUREL. 16, last annotation update)  
 DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LULIBERIN).  
 OS Morone saxatilis (Striped bass).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;  
 OC Moronidae; Morone.  
 OX NCBI\_TaxID=34816;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Chow M.M.; Right K.E.; Gotthelf Y.; Alok D.; Zohar Y.;  
 RA Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 CC EMBL: AF056314; AAD03817.1; -.

DR InterPro: IPR002012; -.  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Amidation; Hormone.  
 SQ SEQUENCE 95 AA; 10411 MW; 980C6988FC279BFC CRC64;

Query Match 85.7%; Score 54; DB 13; Length 95;  
 Best Local Similarity 80.0%; Pred. No. 0.034;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||  
 Db 23 QHMSYGLRPG 32

## RESULT 4

OY1A0 PRELIMINARY; PRT; 99 AA.

AC OY1A0; 01-OCT-2000 (TREMUREL. 15, Created)  
 DT 01-OCT-2000 (TREMUREL. 15, last sequence update)  
 DT 01-MAR-2001 (TREMUREL. 16, last annotation update)  
 DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LULIBERIN).  
 OS Dicentrarchus labrax (European sea bass).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;  
 OC Moronidae; Dicentrarchus.  
 OX NCBI\_TaxID=13489;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Gonzalez-Martinez D., Madiou T., Zmora N., Anglade I., Zanuy S.,  
 RA Zohar Y., Elizur A., Munoz-Cueto J.A., Kah O.;  
 RT "Differential expression of three different prepro-GNRH  
 RT (Gonadotrophin-releasing hormone) messengers in the brain of the  
 RT European sea bass (Dicentrarchus labrax).";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Zmora N., Zohar Y., Elizur A.;  
 RT "3 GNRH form in the seabass Dicentrarchus labrax.";  
 RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY  
 CC -1- SIMILARITY).  
 CC EMBL: AF224279; AAF62898.1; -.  
 DR InterPro: IPR002012; -.  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Amidation; Hormone.  
 SQ SEQUENCE 99 AA; 10758 MW; EC8AEEC93CC02904 CRC64;

Query Match 85.7%; Score 54; DB 13; Length 99;  
 Best Local Similarity 80.0%; Pred. No. 0.035; 1; Indels 0; Gaps 0;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 EHMSTGLRPG 10  
 :|||||  
 Db 27 QHMSYGLRPG 36

## RESULT 5

OY1A0 PRELIMINARY; PRT; 90 AA.

AC OY1A0; 01-OCT-2000 (TREMUREL. 15, Created)  
 DT 01-OCT-2000 (TREMUREL. 15, last sequence update)  
 DT 01-MAR-2001 (TREMUREL. 16, last annotation update)



DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LULIBERIN).  
 OS Rana dybowskii (Frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rana.  
 OX NCBI\_TaxID=71582;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Yoo M.S., Kang H.M., Choi H.S., Chun S.Y., Troskie B., Millar R.P.,  
 RA Kwon H.B.;  
 RT "Molecular Cloning, Distribution and Pharmacological Characterization  
 of a Novel Gonadotropin-Releasing Hormone((Trp)GnRH) in Frog Brain.";  
 RL Mol. Cell. Endocrinol. 0:0-0(2000).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 DR EMBL: AF139911; AAF44343.1; -;  
 DR InterPro: IPR001211; -;  
 DR InterPro: IPR002012; -;  
 DR ProDom: PD000303; -; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Amidation; Hormone.  
 SQ SEQUENCE 90 AA; 10368 MW; C3D573E78B52ABFA CRC64;

Query Match 82.5%; Score 52; DB 13; Length 90;  
 Best Local Similarity 80.0%; Pred. No. 0.071;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EHMSYGLRPG 10  
 :||||| 11  
 Db 25 QHMSYGLSPG 34

RESULT 6  
 Q9DGC8 PRELIMINARY; PRT; 91 AA.  
 AC Q9DGC8;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE PREPRO-GONADOTROPIN-RELEASING HORMONE.  
 GN MDGNRH.  
 OS Oryzias latipes (Medaka fish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Belontiiformes; Adiantichthyidae; Oryziinae; Oryzias.  
 OX NCBI\_TaxID=8090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX PubMed=11006121;  
 RA Okubo K., Amano M., Yoshitake Y., Suetake H., Aida K.;  
 RT "A Novel Form of Gonadotropin-Releasing Hormone in the Medaka, Oryzias  
 latipes";  
 RL Biochem. Biophys. Res. Commun. 276:298-303(2000).  
 DR EMBL: AB041333; BAB16303.1; -;  
 FT CHAIN 22 31 GONADOTROPIN-RELEASING HORMONE.  
 SQ SEQUENCE 91 AA; 10307 MW; A00F2BDE6DE0B5 CRC64;

Query Match 79.4%; Score 50; DB 13; Length 91;  
 Best Local Similarity 70.0%; Pred. No. 0.16;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EHMSYGLRPG 10  
 :|||:| 11  
 Db 22 QHMSYGLSPG 31

RESULT 7

P81749  
 ID P81749 PRELIMINARY; PRT; 10 AA.  
 AC P81749;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE GONADOLIBERIN I (GONADOTROPIN-RELEASING HORMONE I) (GNRH-I)  
 DE (LH-RH) (LULIBERIN I).  
 OS Clupea pallasi (Pacific herring).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Clupeidae;  
 OC Clupeinae; Clupea.  
 OX NCBI\_TaxID=30724;  
 RN [1]  
 RP SEQUENCE AND FUNCTION.  
 RC TISSUE=BRAIN, AND PITUITARY;  
 RX MEDLINE=20114351; PubMed=10650929;  
 RA Carlisle J., Powell J.F.F., Park M., Fischer W.H., Craig A.G.,  
 RA Chang J.P., Rivier J.E., Sherwood N.M.;  
 RT "Primary structure and function of three gonadotropin-releasing  
 hormones, including a novel form, from an ancient teleost, herring.";  
 RL Endocrinology 141:505-512(2000).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 DR InterPro: IPR002012; -;  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Hormone; Amidation; Hypothalamus.  
 FT MOD\_RSS 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RSS 10 10 AMIDATION.  
 SQ SEQUENCE 10 AA; 1105 MW; 284B20B72871F5A3 CRC64;

Query Match 77.8%; Score 49; DB 13; Length 10;  
 Best Local Similarity 70.0%; Pred. No. 0.023;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EHMSYGLRPG 10  
 :|||:| 11  
 Db 1 QHMSYGLSPG 10

RESULT 8  
 P81751 PRELIMINARY; PRT; 10 AA.  
 AC P81751;  
 DT 01-MAR-2001 (TReMBLrel. 16, Created)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)  
 DE GONADOLIBERIN III (GONADOTROPIN-RELEASING HORMONE III) (GNRH-III) (LH-  
 RH III) (LULIBERIN III).  
 OS Clupea pallasi (Pacific herring).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Clupeidae;  
 OC Clupeinae; Clupea.  
 OX NCBI\_TaxID=30724;  
 RN [1]  
 RP SEQUENCE AND FUNCTION.  
 RC TISSUE=BRAIN, AND PITUITARY;  
 RX MEDLINE=20114351; PubMed=10650929;  
 RA Carlisle J., Powell J.F.F., Park M., Fischer W.H., Craig A.G.,  
 RA Chang J.P., Rivier J.E., Sherwood N.M.;  
 RT "Primary structure and function of three gonadotropin-releasing  
 hormones, including a novel form, from an ancient teleost, herring.";  
 RL Endocrinology 141:505-512(2000).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF BOTH LUTEINIZING AND  
 CC FOLLICLE-STIMULATING HORMONES.  
 CC -1- SIMILARITY: BELONGS TO THE GNRH FAMILY.  
 DR InterPro: IPR002012; -;  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Hormone; Amidation; Hypothalamus.  
 SQ SEQUENCE 10 AA; 1230 MW; 284B3233786B45A3 CRC64;

Query Match 74.6%; Score 47; DB 13; Length 10;  
 Best Local Similarity 70.0%; Pred. No. 0.051;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 :|||||  
 Db 1 QHWSYGLRPG 10

RESULT 9  
 ID Q9W7G0 PRELIMINARY; PRT; 33 AA.  
 AC Q9W7G0;  
 DT 01-NOV-1999 (TREMBLERL. 12, Created)  
 DT 01-NOV-1999 (TREMBLERL. 12, Last sequence update)  
 DT 01-MAR-2001 (TREMBLERL. 16, Last annotation update)  
 DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LULIBERIN) (FRAGMENT).  
 GN GNRH2.  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8022;  
 RN [1]  
 RP MEDLINE=99312119; PubMed=10385393;  
 RX Von Schalburg K.R., Sherwood N.M.;  
 RT "Regulation and expression of gonadotropin-releasing hormone gene  
 differs in brain and gonads in rainbow trout.";  
 RL Endocrinology 140:3012-3024(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Von Schalburg K.R., Sherwood N.M.;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 DR EMBL: AF110593; AAD43463.1; -  
 DR InterPro: IPR002012; -  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW AMIDATION; Hormone.  
 FT NON\_TER 33  
 SQ SEQUENCE 33 AA; 3668 MW; 099C825EA7A2A3BB CRC64;

Query Match 74.6%; Score 47; DB 13; Length 33;  
 Best Local Similarity 70.0%; Pred. No. 0.18;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 :|||||  
 Db 24 QHWSYGLRPG 33

RESULT 10  
 ID Q9PT34 PRELIMINARY; PRT; 33 AA.  
 AC Q9PT34;  
 DT 01-MAY-2000 (TREMBLERL. 13, Created)  
 DT 01-MAY-2000 (TREMBLERL. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLERL. 16, Last annotation update)  
 DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LULIBERIN) (FRAGMENT).  
 GN GNRH1.  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8022;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=99312119; PubMed=10385393;  
 RA Von Schalburg K.R., Sherwood N.M.;  
 RT "Regulation and expression of gonadotropin-releasing hormone gene  
 differs in brain and gonads in rainbow trout.";  
 RL Endocrinology 140:3012-3024(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Von Schalburg K.R., Sherwood N.M.;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY  
 SIMILARITY).  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 DR EMBL: AF110593; AAD43461.1; -  
 DR InterPro: IPR001152; -  
 DR InterPro: IPR002012; -  
 DR InterPro: IPR002047; -  
 DR Pfam: PF00446; GNRH; 1.  
 DR PRODOM: PD005116; -; 1.  
 DR PROSITE: PS00256; AKH; UNKNOWN\_1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW AMIDATION; Hormone.  
 FT NON\_TER 33  
 SQ SEQUENCE 33 AA; 3741 MW; 1FE1535E742B7EBB CRC64;

Query Match 74.6%; Score 47; DB 13; Length 33;  
 Best Local Similarity 70.0%; Pred. No. 0.18;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 :|||||  
 Db 24 QHWSYGLRPG 33

RESULT 11  
 ID Q9Z094 PRELIMINARY; PRT; 82 AA.  
 AC Q9Z094;  
 DT 01-NOV-1996 (TREMBLERL. 01, Created)  
 DT 01-NOV-1996 (TREMBLERL. 01, Last sequence update)  
 DT 01-MAR-2001 (TREMBLERL. 16, Last annotation update)  
 DE GONADOLIBERIN PRECURSOR (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LULIBERIN).  
 GN PREPRO-GNRH-1.  
 OS Oncorhynchus nerka (Sockeye salmon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8023;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SPRAIN=NIHKO; TISSUE=BRAIN;  
 RC MEDLINE=96020547; PubMed=8546809;  
 RX Ashihara M., Suzuki M., Kubokawa K., Yoshitura Y., Kobayashi M.,  
 RA Urano A., Aida K.;  
 RT "Two differing precursor genes for the salmon-type gonadotropin-  
 releasing hormone exist in salmonids.";  
 RL J. Mol. Endocrinol. 15:1-9(1995).  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 DR EMBL: D31868; BAA06666.1; -  
 DR InterPro: IPR002012; -  
 DR InterPro: IPR002047; -  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00256; AKH; UNKNOWN\_1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW AMIDATION; Hormone; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 33  
 FT CHAIN 37 82  
 SQ SEQUENCE 82 AA; 9126 MW; C64044EA521B2BBB CRC64;

Query Match 74.6%; Score 47; DB 13; Length 82;  
 Best Local Similarity 70.0%; Pred. No. 0.47;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 :||||| 11  
 DB 24 QHWSYGLRPG 33

## RESULT 12

ID Q9W7G1 PRELIMINARY; PRT; 82 AA.

AC Q9W7G1; 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, last annotation update)  
 DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LUTIBERIN).  
 GN GNRH1.  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8022;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99312119; PubMed=10385393;  
 RA Von Schalburg K.R., Sherwood N.M.;  
 RT "Regulation and expression of gonadotropin-releasing hormone gene  
 differs in brain and gonads in rainbow trout.";  
 RL Endocrinology 140:3012-3024(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Von Schalburg K.R., Sherwood N.M.;  
 RT Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS.  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 DR EMBL: AF110992; A043462.1; -.  
 DR InterPro: IPR002012; -.  
 DR InterPro: IPR002047; -.  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00256; AKH; UNKNOWN\_1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Amidation; Hormone.  
 SO SEQUENCE 82 AA; 9232 MW; 7595B4FCC65FDFD6 CRC64;

Query Match 74.6%; Score 47; DB 13; Length 82;  
 Best Local Similarity 70.0%; Pred. No. 0.47;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 :||||| 11  
 DB 24 QHWSYGLRPG 33

## RESULT 13

ID Q91800 PRELIMINARY; PRT; 82 AA.

AC Q91800; 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, last annotation update)  
 DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LUTIBERIN).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8022;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;  
 RA Ferriere F., Bailhache T., Jégo P.;  
 RT "Oncorhynchus mykiss scNRH-1 cDNA from brain";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 DR EMBL: AF232212; AAF91280.1; -.  
 DR InterPro: IPR002012; -.  
 DR InterPro: IPR002047; -.  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00256; AKH; UNKNOWN\_1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Amidation; Hormone.  
 SO SEQUENCE 82 AA; 9198 MW; 7595A0B896556A69 CRC64;

OY 1 EHWSYGLRPG 10  
 :||||| 11  
 DB 24 QHWSYGLRPG 33

## RESULT 14

ID Q918P9 PRELIMINARY; PRT; 82 AA.

AC Q918P9; 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, last annotation update)  
 DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LUTIBERIN).  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8022;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ferriere F., Bailhache T., Jégo P.;  
 RT "Oncorhynchus mykiss scNRH-1 cDNA in the brain";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 DR EMBL: AF232213; AAF91281.1; -.  
 DR InterPro: IPR002012; -.  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Amidation; Hormone.  
 SO SEQUENCE 82 AA; 9203 MW; 8053F4F221A0FF08 CRC64;

Query Match 74.6%; Score 47; DB 13; Length 82;  
 Best Local Similarity 70.0%; Pred. No. 0.47;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 EHWSYGLRPG 10  
 :||||| 11  
 DB 24 QHWSYGLRPG 33

## RESULT 15

ID Q9PSY9 PRELIMINARY; PRT; 88 AA.

AC Q9PSY9; 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)  
 DT 01-MAR-2001 (TREMblrel. 16, last annotation update)

DE GONADOLIBERIN (GONADOTROPIN-RELEASING HORMONE) (GNRH) (LH-RH)  
 DE (LULIBERIN) (FRAGMENT).  
 OS Sparus aurata (gilthead sea bream).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;  
 OC Sparidae; Sparus.  
 OC NCBI\_TaxID=8175;  
 OX  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Ovary;  
 RA Nabliasi M.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: STIMULATES THE SECRETION OF GONADOTROPINS (BY  
 CC SIMILARITY) TO THE GNRH FAMILY.  
 CC -1- SIMILARITY: TO THE GNRH FAMILY.  
 CC EMBL; AF046799; AAD02425.1; -.  
 DR InterPro: IPR002012; -.  
 DR Pfam: PF00446; GNRH; 1.  
 DR PROSITE: PS00473; GNRH; 1.  
 KW Amidation; Hormone.  
 KW NON\_TER 88  
 FT SEQUENCE 88 AA: 9788 MW: F7EB868C2FBD19F CRC64;  
 SQ

Query Match 74.68; Score 47; DB 13; Length 88;  
 Best Local Similarity 70.08; Pred. No 0.51; 2; Indels 0;  
 Matches 7; Conservative 1; Mismatches 2; Gaps 0;

OY 1 EHWSTGLRPG 10  
 : ||||| ||  
 DB 24 QHWSYGMPLPG 33

Search completed: November 13, 2001, 14:34:00  
 Job time: 426 sec